3-26-2015

Entity Recognition via Multimodal Sensor Fusion with Smart Phones

John E. Nagy

ENTITY RECOGNITION VIA MULTIMODAL SENSOR FUSION WITH SMART PHONES

THESIS

John E. Nagy, Capt, USAF

AFIT-ENG-MS-15-M-023

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENG-MS-15-M-023

ENTITY RECOGNITION VIA MULTIMODAL SENSOR FUSION WITH SMART PHONES

THESIS

Presented to the Faculty
Department of Electrical and Computer Enginnering
Graduate School of Engineering and Management
Air Force Institute of Technology
Air University
Air Education and Training Command
in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Computer Science

John E. Nagy, B.S.C.S.
Capt, USAF

March-2015

AFIT-ENG-MS-15-M-023

ENTITY RECOGNITION VIA MULTIMODAL SENSOR FUSION WITH SMART PHONES

THESIS

John E. Nagy, B.S.C.S.
Capt, USAF

Committee Membership:

Kenneth M. Hopkinson, PhD
Chair

Meir Pachter, PhD
Member

Maj Brian G. Woolley
Member

AFIT-ENG-MS-15-M-023

Abstract

This thesis serves as an exploration that takes the sensors within a cell phone beyond the current state of recognition activities. Current state of the art sensor recognition processes tend to focus on recognizing user activity. Utilizing the same sensors available for user activity classification, this thesis validates the ability to gather data about entities separate from the user carrying the smart phone. Two Experiments of exploring different sensing techniques are performed to determine the ability to classify entities with smart phone sensor data. The first experiment focuses on classifying stationary entities affecting the environment near a smart phone. The second experiment focuses on classifying an automotive entity moving past a smart phone. Using statistical and wavelet attributes for classifying the entities in the two experiments, respectively, it is possible to accurately classify entities based off the entities environmental influence. With the ability to sense entities, the ability to recognize and classify a multitude of items, situations, and phenomena opens a new realm of possibilities for how devices perceive and react to their environment.

Table of Contents

Page

List of Figures

List of Tables

ENTITY RECOGNITION VIA MULTIMODAL SENSOR FUSION WITH SMART PHONES

I. Background

Over the past two decades, cell phones have exploded into a nearly ubiquitous presence in society. The penetrative extent of cell phone use is felt in not just industrialized nations, but also developing nations. The cell phone has become an equalizer of sorts, helping all people with access to cell based communications enjoy a sort of homogeneity of communications access. This has allowed farmers and craftsmen in far flung corners of the world to communicate and gain access to the global commerce system, not to mention the familial and communal benefits inherent in better communication.

The introduction of the smart phone ushered in yet another explosion in capability. No longer was the personal computer a monolithic item that was expensive in not just terms of currency, but resource and space requirements as well. Smart phones put the power of a computer into the palm of many more hands than would have been possible otherwise. Coupled with the communication infrastructure required for cell-based communications, smart phones offered numerous additional benefits. These include, but are not limited to, enhanced communications through social applications (apps), search capabilities to more easily seek out global connectivity, access to medical care advice, notification of pending disasters (both natural and manmade), and general information accessibility for everything from sports to crop planting.

The continued evolution of smart phones via increased computation powers and communications bandwidth will ultimately narrow the gap further between smart phones and computers. Additionally, the inclusion of multimodal sensors within smart phones gives them unique abilities unavailable to traditional computing platforms. The American Heritage College dictionary defines **modal** to be "of, relating to, or characteristic of a mode", as such a multimodal sensor package would be the combination of more than one sensor capable of sensing different characteristics [30]. Multimodal sensors present in smart phones include, but are not limited to, accelerometers, magnetometers, gyroscopes, microphones, thermometers, and barometers. The sensing capability present in an average smart phone is far beyond the sensing capability present in an average personal computer.

The inclusion of multimodal sensors within a smart phone allow the phone to be used in manners not possible in the personal computing revolution. In addition to offering many of the benefits of a personal computer, the smart phone's ability to sense the environment allows for: the tracking of activities, the detection of entities external to the smart phone, and environmental surveillance. The sensing ability presented by the inclusion of multimodal sensors opens the door to a wide range of possibilities, with future additions to the sensing suite offering further expansion in what can be sensed.

The ability to detect, interrogate, and classify phenomena sensed by a smart phone is wholly dependent on having versatile and intelligent written software paired with the sensors on the smart phone. Prior to the development of smart phone based activity recognition, software developers and computer scientists had been using specialized sensing devises to detect user activity and environmental conditions. With the addition of the first solid state accelerometer, software developers and computer scientists quickly set to work on methods to attempt to detect user and environment conditions with smart phones [2, 4]. The addition of gyroscopes and magnetometers led those developing recognition programs down an ever increasing work of discovery. Quickly the science evolved from merely recognizing the current position of a smart phone to recognizing, the location of a smart phone relative to a person, specific transportation modes, as well as environmental phenomena [11, 12, 17].

An area yet to be explored in detail is the ability to look beyond the smart phone and determine whether the sensor data gathered by the smart phone can ascertain the presence of devices, environmental phenomena, and other entities. With the ability to utilize sensor data to determine phone orientation, magnetic heading, activity recognition, and more, there is plenty of research available to push the sensing capability further. Entity recognition via smart phone sensors utilizes prior research based on accurate activity recognition. Attribute selection, classification algorithm generation, and axis synthesis techniques are combined to process sensor data and utilize the data in recognizing entities that affect the smart phone's environment [18, 44, 35, 13]. A smart phone, with its increasing array of embedded sensors capable of sensing a diverse set of environmental attributes, is capable of sensing non-smart phone centric entities. Smart phones, with their suite of sensors, are capable of recognizing entities that affect the environment in ways detectable by the smart phone's sensors.

In the detection of phenomena via a multimodal suite of smart phone sensors, it is apparent that the current literature comes short when describing what is being detected. Using the term *activity* to describe a user's physical motion and/or the transportation mode being utilized is accurate enough [18, 44]. Using *activity* to describe the activity acting on an environment a smart phone is monitoring, such as an earthquake, becomes less understandable, because while the phone is experiencing a shaking activity, the entity causing the shaking is the earthquake. In this research, the term **entity** describes environmental actors being measured by the smart phone sensors. The American Heritage College Dictionary defines **entity** to be "the existence of something considered apart from its properties," as such the goal of this research is to determine whether there is validity to the use of a multi-modal approach to entity detection [30].

Through systemic experimentation, the author proves there is validity to using a smart phone sensor suite to detect and recognize entities acting on the environment in manners observable by the smart phone sensors. The first experiment evaluates the ability of a smart phone's sensors to detect the environmental attributes affected by entities. The environmental attributes are captured via their respective sensors and then processed through various decision trees, proving the ability to accurately classify between different but similar entities. The second experiment evaluates the ability of a smart phone to be used as an environmental scanner to detect the passing of a specific entity. With the smart phone in a stationary position, an entity that generates a magnetic signature is passed over the phone and the ability of the sensor to capture data for recognition purposes is validated.

As hardware engineers construct cell phones with an increasing number of sensors, capable of sensing unique and varied environmental attributes, the developer and scientist are faced with the challenge of combining the data streams from multiple sensors to ascertain an environmental attribute. Specific combinations of sensor data streams can be combined to detect certain user or environmental attributes. For instance, it is possible to determine whether a user is sitting or climbing stairs via the sensors in their cell phone [1, 37, 25, 3, 34]. In the environment, Faulkner has shown it is possible to detect earthquakes with a cell phone [12, 11]. Through the experiments detailed within, the author proves it is possible to recognize a large and diverse set of entities with smart phone sensor data.

Traditionally, user activity has focused on the actions of a single entity. Research has been performed to identify the smart phone sensors most able to accurately classify which activity a user is performing. Algorithms have been developed to recognize whether a user is walking, jogging, climbing or descending stairs, biking, sitting, standing, taking off or landing in an airplane [41, 35, 1, 6]. The collection of user locations via signal triangulation and/or GPS location analysis allows for the reporting of traffic conditions [33]. By utilizing cell phone data from more than one user, developers and scientists have shown the ability to use the location data to determine the level of congestion of road and highway systems, thus producing an awareness of a system's status without requiring live video feeds. The ability to collect and aggregate the data from cell phones opens the possibility for even greater environmental awareness.

Scientists have delved beyond the task of activity classification and identification; using the community of smart phones present in the population, they have begun to aggregate data from multiple sensors to identify environmental attributes. The work of Faulkner has shown that it is possible to use aggregate data from a community of cell phone sensors to determine whether and where an earthquake has occurred [12, 11]. This aggregation of accelerometer and gyroscopic data shows that it is possible to acquire real world conditions from a smart phone's sensor suite. It is intuitive to believe that there are a multitude of real world conditions that can be acquired, analyzed, classified, and identified with varying degrees of accuracy by aggregating the sensor data from a smart phone.

The terms multimodal and sensor fusion have both been used to describe the combination of multiple sensor output data streams into a product that can determine whether a condition is being met. Multimodal sensing is the utilization of more than one sensor and sensor fusion is the utilization of output from more than one sensor. Typical smart phone sensor fusion occurs in relation to the various algorithms that are used to determine the orientation and/or movement of a device. The orientation and movement algorithms utilize the output of a cell phone's magnetometer, accelerometer, and gyroscope to determine how a device is oriented. Additionally, they utilize the sensor outputs to determine whether the device is experiencing tilt or sharp directional changes. These determinations (gathered from the outputs of the sensors) are then used as inputs in various software applications to aid in game play or measurements (i.e. digital levels and compasses). As no single signal from either the magnetometer, accelerometer, or gyroscope is able to definitively identify whether a device is moving or oriented in a specific direction, it is the aggregation, or fusion, of multiple sensor output signals that are input into an algorithm to determine whether a condition exists.

The aggregation of smart phone sensor data has value in determining the environmental status affecting a smart phone. Through the careful analysis of aggregate smart phone data, it may be possible to determine a myriad of conditions present in the environment. Events beyond traffic congestion analysis are possible with aggregate smart phone sensor data. With the proper classification algorithms, it should be possible to determine what modes of transportation a user is utilizing; whether a triathlon has just started and which leg the participants are taking part in; whether users are at a concert or evacuating a building due to some emergent situation, what appliance or machinery is present in a cell phone's immediate environment; and possibly even whether a large geomagnetic storm is taking place [18]. Using the processes developed by scientists for activity recognition, the author collects, aggregates, and processes the multimodal smart phone sensor data to accurately classify entities. This accurate classification proves that the basis for entity recognition is possible with modern techniques.

The presence of accelerometers that can measure changes in the force of gravity accurately to 0.001, gyroscopes that can measure changes in inertia in degree per second to 0.001 of a degree, and magnetometers that can measure changes to the magnetic field down to a resolution of 0.000001 tesla presents the opportunity for the replacement of certain legacy sensors with a smart phone deployed to monitor, analyze, and record certain events [39, 40, 7].

While the near ubiquity of smart phones and the suite of sensors they contain make for an attractive research target, the ability to recognize external, non-transportation entities is an area of study still very much in the early stages. Most research that fits the external, non-transportation entity detection and recognition has been limited to the large-scale natural events such as earthquakes [12, 11]. There is reason to believe the algorithms developed for the analysis of a smart phone's environment can be utilized in more specialized situations. The introduction of gyroscopic, acceleration, and magnetic detection sensors into the vehicle and/or uniforms of military personnel and first responders could prove useful to the detection of a number of unique conditions whose existence would alert an incident or combat command center to the presence of a condition that requires immediate attention. Thus, the aggregation of smart phone sensor data could have implications far beyond the life of everyday phone users.

The ability to capture recorded data from an entity, whether it be an object or an event, from a single user is useful, but to truly get a grasp on the magnitude of an entity, it would be necessary to get the data from as many sensors (cell phones) as possible. Thus the ability to detect an event and send a capture request to all smart phones within a given radius may be useful for detecting certain environmental actors. This line of research has been explored in regards to large scale phenomena like earthquakes, and civilian crowd movements classified as flocks [12, 11, 24]. Combined with the ability to detect additional entities, the research focused on detecting large scale events offers intriguing possibilities.

A review of the literature regarding the evolution of smart phone sensor utilization is necessary to understand how sensor utilization has changed. An understanding of where scientists have taken the art of sensing since before smart phones to where we are today helps reveal the nuanced techniques used to coax the most accurate sensor data into algorithms for device orientation and user activity identification. The more complex aspects of recognition such as the classification algorithms and attributes utilized are varied, however, so are the more simplistic aspects such as the computations used to stabilize a smart phone's orientation.

II. Literature Review

2.1 Smart Phones

The literature surrounding the use of smart phones as sensing platforms has exploded over the past decade and shows no sign of slowing down. In order to get a grasp on the state of research surrounding multimodal sensor fusion it is best to have an idea on how the field has evolved. Field evolution has been guided (initially) by the presence of a limited number of sensors in the cell phone. As more sensors have been added, scientists and developers have produced research to utilize the additional sensors to increase the ability of a smart phone to determine user activity and increase environmental awareness. The earliest smart phone sensor programs relied on using the WiFi and baseband capabilities of the cell phone [33] [20]. Over time, cell phone manufacturers added accelerometers, magnetometers, GPS, gyroscopes, and barometers to their phones. As tends to happen, researchers have taken advantage of the additional capabilities offered by the smart phones and developed ever more complicated packages to gauge and track user and and environmental activity. In addition to tapping into the latent sensing abilities offered by smart phones, researchers have to be mindful of the features being selected to obtain information regarding a task. Thus algorithms have been developed that seek to balance the tracking ability offered by a smart phone with resource preservation and user interaction (or lack of interaction in the case of training data). This review of research literature will capture the origins of sensor fusion, feature selection activities, algorithm development, and resource utilization. Utilizing the knowledge obtained via the prior research aids in taking the next step towards entity recognition.

2.2 Sensor Fusion

The first place to start would be the why of multimodal sensor fusion. Before determining whether there is value to fusing the sensor data that is output by multiple sensors, one must acknowledge that value exists in the act of abstracting data from sensors in the first place. While the intuition is almost certainly that there is value to processing sensor data, the scope and depth of mining extends far beyond what most of the populace considers possible. In matters of scope, the sensor can be viewed first and foremost as providing a 'status' on the 'state' of a smart phone. Between the sensors mentioned above and additional sensors in the phone such as proximity and battery temperature sensors, the first order of business is to provide a status to the phone. It is through such status reporting that an algorithm in the phone determines when a phone has been lifted to an ear to make a phone call, thus turning the screen off, or that a phone left in the sun is getting dangerously hot, thus shutting the phone off. The use of accelerometers has been used in devices with a hard disk drive to park the drive heads when certain gravity thresholds are violated, thus protecting the data on the drive.

Moving beyond contributing to the phone's basic operations, accelerometers were introduced as a means to detect orientation. With display screens that can rotate beyond a landscape and portrait display, the ability to recognize how a phone was oriented proved useful. This sensor has been seized upon by game makers, as well as those interested in detecting activity, as a means to detect what a user is doing with their phone. And while the accelerometer is useful for detecting between a portrait or landscape orientation of the phone, the inclusion of powerful 3-axis accelerometers combined with 3-axis magnetometers allowed for fine grain measurements. By combining the data from an accelerometer and magnetometer, the cell phone could now determine not just orientation, but via the devices magnetic orientation algorithms can determine the pitch, yaw, and roll of a device. This allows the device to be used for recording, recognizing, and responding to very specific movement scenarios. The inclusion of a gyroscope and barometer have allowed for even finer grain activity detection, allowing for the accurate detection of turns and altitude change, respectively [2].

Moving beyond the device and to the user, there are numerous studies and programs that have been developed that determine the activities of a user. Research has been done to determine whether a user is stationary or moving, whether a user is standing or sitting, whether a pedestrian is walking or biking, whether a pedestrian is walking or running, whether a user is moving via pedestrian means or motorized transport, whether a user is in a bus or a subway, and so on. The idea of determining user activity has merits from the concept of activity classification for logging purposes to the analysis of travel patterns for transportation system development and tuning. By adding GPS chips to the cell phone, the user is not just provided awareness of their longitude and latitude, but applications granted access to location data and analyze travel information for traffic congestion reporting, emergency service location reporting, and even opening the possibility for disease tracking and reporting [31]. The scope of usefulness to cell phone sensor data has moved far beyond the earliest iterations where they were useful to not much more than the hardware and software of a single cell phone user.

The inclusion of baseband and WiFi chipsets in a smart phone is ubiquitous in so much that for a phone to be a smart phone, it will include a chipset to access a communications network as well as the ability to connect to public and private local wireless networks. Poolsawat, Pattara-Atikom, and Ngamwongwattana discuss fusing base transceiver station (BTS) information with GPS location data for providing status on traffic [33]. In attempting to implement an alternative to the system of surveillance cameras and sensors that local transportation departments install to monitor traffic conditions, Poolsawat et al. focused on cost, ease of deployment, and systemic robustness. The costs to building and operating an effective traffic surveillance system are high. The role of a traffic information system is to monitor traffic conditions, process the conditions, and broadcast solutions to certain conditions. This set of traffic monitoring tasks is ideally suited to a hybrid system that combines some non-cell based sensors and data acquired from user's cell phones. Poolsawat et al. detail a system that captures data from the endpoint (a cell phone user); the endpoint interacts with the cellular provider's BTS and it is this interaction that proves useful to traffic monitoring. Using software to abstract data features from the cell phone's BTS interaction, Poolsawat et al. build a system that indicates the mobile country code, the mobile network code, the location area code, and the cell ID (CID) of the BTS a cell phone is currently associated. Using the data features abstracted from the BTS information, the authors calculate a cell dwell time (CDT) whereby the duration of time a endpoint cell phone spends within a particular cell ID is identified and sent to a collection server for analysis. Using a history of CDT in each CID, analysis can be performed to determine whether a user is in a congested traffic zone. Adding GPS coordinates to the traffic system data would enable for precise identification of congestion points, however, due to GPS receivers requiring line of site connectivity to the GPS satellites, there is no guarantee. Whereas, if a cell phone can communicate with a BTS, it will always have a CID to reference. In addition to preserving privacy by not sharing user details, the system can be configured to preserve system resources such as bandwidth and battery life; in order to preserve resources, the system would not maintain a constant state of connectivity between end user and server, the server would receive a feature report once every 3 minutes or so.

Using accelerometer data from a user's cell phone can produce a trove of feature data by which to evaluate numerous activities as well as potential environmental attributes. However, the data output by the accelerometer would be useless if it cannot be oriented relative to gravity. As an accelerometer measures the strength of gravity, it is possible to determine the orientation of an accelerometer (and thus the device housing the accelerometer) relative to gravity. In the article, Using Gravity to Estimate Accelerometer Orientation, David Mizell articulated a methodology to determine device orientation with a three-axis accelerometer [32]. By using an estimate achieved by averaging the accelerometer samples, the gravity constant can be determined. Letting **v** represent the average of acceleration for a given time interval window, we have:

$$v = (v_x, v_y, v_z)$$

Let **a** represent a point of time within the window, we have:

$$a = (a_x, a_y, a_z)$$

Using the average **v** and the instantaneous **a** it is possible to calculate both the static and dynamic acceleration experienced by the accelerometer. The static acceleration corresponds to the effect of gravity and the dynamic acceleration corresponds to the effect a user's activity has on the accelerometer. Letting **d** represent the dynamic component of **a** we find:

$$d = (a_x - v_x, a_y - v_y, a_z, v_z)$$

The orientation of the device can then be found by using the vector dot products. This is done by computing the projection **p** of **d** upon the vertical axis **v** as:

$$p = \frac{d \cdot v}{v \cdot v}$$

Whereby **p** is the vertical component of the dynamic acceleration vector **d**. From the vector **p** we can compute the horizontal component of the dynamic acceleration:

$$h = d - p$$

Through the above equations it is possible to decompose the accelerometer readings to obtain the gravity manifested upon the accelerometer, thus allowing the orientation of the device to be calculated. While the intent of Mitzel's work was to prove that device orientation could be determined by transforming accelerometer data, thus focusing primarily on the vertical gravitational component, later work proved that horizontal movement could reliably be determined once device orientation had been calculated [19].

Though much of the prior research utilizes the multimodal sensing, it is most often performed in a complimentary manner to enhance the measurements obtained with one sensor with added data from another [2]. In other instances a multimodal approach is used in a supplementary manner in case one sensor fails to perform as expected [33]. The goal of this research is to determine whether an entity can be accurately classified using both a complimentary and supplementary method by building classifiers based off all available data.

2.3 Sampling Windows

As noted in [32], in order to achieve an average of acceleration for $\mathbf{v}$, it is necessary to designate a window length. Across the literature there are examples of various window lengths, with [32] indicating a length of a few seconds to others indicating windows of up to eight seconds. In research performed in Activity Recognition on an Accelerometer Embedded Mobile Phone with Varying Positions and Orientations, the window length was found to be between the four and five second time frames [41]. The goal of the research was to refine the science of activity recognition with an accelerometer, regardless of the position and orientation a user has their cell phone. Earlier work cited by Sun, Zhang, Li, Guo, and Li had devised methods to extract features that could be used to identify various types of pedestrian activity, but were limited in that they required the user to mount the sensor to their body in a specific location orientation. Through various algorithm refinements, the Sun et al. present a method to free users from such stringent orientation requirements for accurate activity detection. In developing their orientation insensitive technique, Sun et al. propose using the magnitude of the accelerometer readings to compensate for changes in device orientation. As such, Sun et al. generate an additional feature from the accelerometer output:

$$(A, \|A\|) = (a_x, a_y, a_z, \|a_x, a_y, a_z\|)$$

With the orientation insensitive feature, Sun et al. found that using an overlapping window divided into frames, which was able to accurately recognize activity 93.1% of the time when the window length was set to 4 seconds. Using a orientation sensitive methodology increased the window to 5 seconds; this demonstrates that depending on the computational methods used, window length will vary. In both cases, the windows were divided into frames of 1 second duration for feature extraction.

The windows necessary for activity and entity recognition are different, as accurate classification of an activity can be thought of as recognizing a pattern that takes place over a relatively long duration of time compared to the recognition of an entity that may be affecting the environment around a smart phone for a brief duration. Thus a snapshot of an activity pattern will contain the information necessary to determine the activity being performed, whereas a snapshot of an entity could be during any number of potential patterns depending on the effects being generated by the entity. Optimally, a window would capture an entire cycle of a mode of operation for an entity, eliminating the need to classify multiple operation phases. This research will show that utilization of entity signature snapshots results in accurate classification of the entities used as control variables.

In addition to identifying optimal window length, the Sun et al. sought to identify means to achieving accurate activity recognition while preserving cell phone resources [41]. After determining which sensors will be used in an activity recognition task, the next task is to determine sampling rates and feature selection. Sampling rates will vary greatly from activity to activity. When sampling for human activity recognition, relatively lower sampling rate of 20 - 60Hz have proven sufficient. When sampling non-human activity recognition, higher rates of sampling may be required, that is one of the tasks of this research. In either case, a higher sampling rate may not be resource conservative, but it will provide data for analysis. In regards to feature selection, Sun et al. generate the following for each frame: the mean, variance, Frequency-Domain entropy, FFT energy, and the correlation. When selecting features, Sun et al. tended to select less computationally complex features to save resources. For recognition purpose, each activity being recognized will have features for the mean, variance, Frequency-Domain entropy, and FFT energy. As such, for a system capable of recognizing standing, walking, biking, and running, there would be 16 features to train against plus 6 more for correlation between the 4 activities. Using an extracted feature vector, Sun et al. normalize each extracted feature vector before training.

2.4 Orientation and Position

Normalization, or transformation, of a signal is important when discussing generation of features from a sensor. As noted prior, compensating for the orientation of a users cell phone is important. One method involves calculating the static gravitational component to determine which axis is vertical [32] and an alternate method involves taking the magnitude of the each accelerometer component to compensate for device orientation [41]. A third technique utilizes the concept on weightlessness; an accelerometer will experience weightlessness when carried on a person that is running or jumping, thus revealing the vertical axis of the accelerometer [19]. In either case, when generating features from sensor data it is necessary to transform the data. Sensor data or signals are transformed into a common coordinate system in order to improve activity recognition. As noted in Accurate Activity Recognition Using a Mobile Phone Regardless of Device Orientation of Location, device orientation is not the only concern [19]. The location of the cell phone is also vital to activity recognition as user movement will look significantly different to cell phone sensors depending on where the device is held; a sensor at the waist will experience different force signatures then a device strapped to the upper arm. In order to compensate for location variation, Henpraserttae, Thiemjarus, and Marukatat devise a method of feature training that involves different feature signatures for each body position. As an example, the feature set for recognizing whether a person is running with a cell phone on their arm will exhibit different characteristics from a person is running with a cell phone in a pocket. Using robust features sets that are independent to device placement is done by creating a model specific to each likely area of device placement.

Henpraserttae et al. [19] explore methods to calculate the forward axis. Adding to the work performed by Mizell [32], they utilize the mean of the dynamic acceleration experienced by the accelerometer. They assign the dynamic portion of the vertical axis to **w** and use it to find the forward axis. Under the assumption that most activity is in the forward-backward direction, the forward direction can be computed from the principal axis of data on the plane that is perpendicular to **w**:

$$x'_t = x_t - (x_t^T w)w$$

where x' is the removed acceleration signal along the vertical axis and x is the raw accelerometer signal. Next, an eigen-decomposition is performed on the covariance matrix of the projected data:

$$C = \frac{1}{T}\sum_{t=1}^{T}(x'_t - \mu')(x'_t - \mu')^T$$

where μ' is the mean of the projected data, calculated by:

$$\mu' = \frac{1}{T}\sum_{t=1}^{T} x'_t$$

The forward axis is parallel to the main eigenvector of the covariance matrix **C**. With **u** corresponding to the eigenvector that has the largest eigenvalue, **u** will be used as the forward axis in Henpraserttae's et al. global coordinate system.

$$\text{if } x_1'^T u < 0 \text{ then } u = -u$$

Knowing **u** to be the horizontal axis and **w** to be the vertical axis, the last axis to find is the sideward axis:

$$v = u \times \mu$$

Knowing the three axes, one can construct the transformation matrix as:

$$T = \begin{bmatrix} u_x & u_y & u_z \\ v_x & v_y & v_z \\ w_x & w_y & w_z \end{bmatrix}$$

Having established the matrix, Henpraserttae et al. use the dynamic mean to estimate the rotational angles for when the device is placed in different orientations. The rotational matrix is used to transform the input signal into the same reference coordinate system regardless of orientation or placement. Thus for activity recognition purposes, the first task is to classify the probable location then to classify the activity taking place by comparing the normalized values to training datasets. Henpraserttae et al. found significant differences between classification without and with transformation, with transformative accuracy performing better by $42-51\%$ in training sets with a minimal number of orientation classifications. When more orientation classifications are trained on, the accuracy for activity recognition is 5.8% higher than without classification. In all cases, a normalized classification system outperformed using non-transformed data.

2.5 Multimodal Data Fusion and Information Presentation

Moving beyond orientation and position, researchers continue to identify the most useful sensor data streams to fuse when capturing data for later analysis playback. Previously discussed research has utilized BTS and GPS data to identify a cell phone's location. A good deal of activity recognition software utilizes GPS data in addition to the accelerometer, though for pedestrian activity recognition purposes the GPS data is most often treated as a perk rather than a means to detecting and classifying a user's activity. Research performed by Microsoft fused the use of camera, accelerometer, gyroscope, magnetometer and GPS data [4]. Researchers designed the Greenfield program as a demonstration application to help smart phone users locate their cars, though the breadcrumb left by the phone could be used to locate any number of entities. Through the use of accelerometer features, the program counted the user steps. Gyroscope features helped identify turns through changes in inertia. Magnetometer features identified compass bearing, though external interference from building structures and items in pockets and purses limit the usefulness of the magnetometer for determining *true* compass bearing. GPS location data was available in non-parking garage scenarios, but once a vehicle was parked in a covered location, GPS data became unreliable. The camera was used to capture the exact state and location a vehicle was parked in. Together, the researchers used this data to create a breadcrumb trail where users could walk back to their vehicle with bearing, step counts, and turn instructions. Besides providing an integrated use of fused data, specifically the breadcrumb trail generated by accelerometer, gyroscope, and magnetometer input, the researchers also studied the cognitive effects the data presentation would require of users. As a data presentation application, users found Greenfield presented information that may have been highly accurate but was mentally taxing to process.

Continuing to develop the concept of activity recognition with cell phone sensors is important for the purposes of physical activity monitoring, personal impact and/or exposure monitoring, and transportation and mobility-based recruitment. In an effort to distinguish between pedestrian mobility and vehicular mobility, the Reddy et al. of Using Mobile Phones to Determine Transportation Modes developed fine grain activity recognizers that worked independent of external knowledge [35]. Previous full-featured activity recognizers used external indexes to identify likely transportation hubs; Reddy et al. rely more heavily on a combination of GPS and accelerometer data to identify mass transit. As GPS is found to perform satisfactorily when attempting course grained transportation mode classification, and then only when signals are present, trying to classify systems with similar speed and acceleration profiles requires finer grained signature classification. The accelerometer in the iPhone 5, for instance, is able to measure gravity with an accuracy of 4 milli-gravity [39]. Using the accelerometer data, Reddy et al. produce accurate acceleration and breaking signatures for transportation modes that present similar speed profiles. Through the fine-grained accelerometer data, classification between buses, trains, and subways is more accurately determined. In addition, through techniques developed both previously and new introductions in their research, Reddy et al. produce a more robust solution that is device, location, and orientation agnostic.

Investigation into sensing techniques beyond GPS and accelerometers was researched in [35]. Reddy et al. researched using wireless infrastructure recognition to obtain accurate transportation classification results. The research indicated that the wireless technology such as bluetooth is not pervasive enough, and WiFi and BTS are too dependent on a dense distribution and are not suitable for fine-grain details. A combination of GPS and accelerometer data was found to produce the most accurate classification while preserving system resources. GPS data features proved useful for determining activity due to speed distribution (range). Accelerometer data features proved useful for determining variance of motion changes. Some examples where a combination of the two sensor data features prove useful for discrimination are when differentiating between walking and running and biking. Walking and running may exhibit similar speed characteristics based off of GPS data, but the variance in accelerometer output will be larger when running; the same traits are exhibited when comparing running and biking, with similar speed characteristics being possible and running having more accelerometer variance than biking. With all three activities able to take place at the same location, referencing an external database of transportation modes would not offer much fidelity in accurate activity recognition. However, by using GPS to determine speed, and using the data features extracted from the accelerometer output, accurate recognition of an activity is increased. In regards to resource preservation, as the use of the GPS sensor is more resource intensive then the use of the accelerometer, the GPS can be left off when the accelerometer is not detecting any motion.

2.6 Normalization and Classification

Reddy et al. [35] found that their techniques allows for a data window size of one second, allowing for a 75% reduction from the results presented in [41]. The authors found that a one second window allowed for near instantaneous classification of transportation mode. Smaller window size resulted in inaccurate activity recognition, larger window size results in unnecessary delay in activity recognition. Accelerometer data was normalized by taking the magnitude of the readings:

$$A_{mag} = \sqrt{(A_x)^2 + (A_y)^2 + (A_z)^2}$$

which allows for the assumption of random and possibly changing device orientation. Features from the accelerometer data are the mean, variance, energy Discrete Fournier Transform (DFT) coefficients. From the GPS data, the feature utilized was speed with the algorithm weeding out invalid points. Activity recognition and classification was done with correlation based feature selection (CFS). CFS was chosen because it allowed for a feature subset selector that eliminates irrelevant and redundant attributes. Examples of the utilization of the various features are: GPS feature used to differentiate between still and motorized transport, accelerometer variance used to determine whether an individual is running, and accelerometer DFT data used to differentiate between different on-foot transportation modes.

In addition to selecting the features most relevant to activity recognition from the sensor data set, Reddy et al. explored which classification system selected the correct activity [35]. The instance classifiers considered by Reddy et al. were the C4.5 Decision Trees (DT), K-Means Clustering (KMC), Naives Bayes (NB), Nearest Neighbor (NN), and Support Vector Machines (SVM). Additionally, a continuous Hidden Markov Model (CHMM) and a two-stage system involving the most accurate instance based classifier (the C4.5 DT) combined with a discrete Hidden Markov Model (DHMM). The classification structure functioned as follows:

Data → Noise Filtering → Feature Calculation → DT Instance Based Classifier →

→ DHMM Classifier

which would select the transportation mode classification. With the above classification structure in place, the research looked at the accuracy as it related to device placement. When the device was carried in the hand or mounted to the upper arm, the accuracy was the highest; waist, pocket, bag, and chest placement resulted in lower accuracy ratings, though the accuracy between the lowest and highest rates were between 94.3% and 95.0%. Some of this lack of precision can be made up for with user specific training. With user specific training, the accuracy increased 2.2% as compared to the generalized classifier. Overall, this study produced highly accurate activity classification across both pedestrian and motorized methods with utilizing energy aware detection to minimize resource strain without requiring user specific training or external indexes. Lastly it showed that accurate prediction could be achieved through location and orientation agnostic processes.

Expanding on which features have value when used to discriminate between activity types, Anjum and Ilyas [1] seek through techniques similar to Reddy et al. [35] to determine the most accurate data features to extract. The research performed in this paper was limited to classifying pedestrian means of transportation (plus driving); recognizable activities were walking, running, ascending stairs, descending stairs, cycling, driving, and being inactive. As in [35], a number of instance classifiers were examined with the C4.5 DT proving the most reliable. As a multi-modal experiment, data streams from the accelerometer (3-axis), gyroscope (3-axis), and the GPS (latitude, longitude, and altitude) were acquired. Anjum and Ilyas researched the optimal sample rate for acquiring data to classify and found that 8Hz proved adequate for human activity; sampling rates from 5Hz to 100Hz were investigated with 8Hz proving the optimal rate. As noted in previous studies, the varying orientation of a phone does not allow for a meaningful comparison of measurements of a particular axis' data with the measurements from the same axis in a different activity trace. Anjum and Ilyas chose to use a Eigen-decomposition of the covariance matrix for the 3 accelerometer axis in order to rotate the three orthogonal reference axes $d_1, d_2,$ and d_3. The three orthogonal axes are organized to the axes descending order of signal variation. In preprocessing, the sample covariance of any two axis i and j is computed via:

$$\sigma_{ij} = \frac{1}{N-1}\sum_{n=1}^{N}(a_i[n]-\bar{a}_i)(a_j[n]-\bar{a}_j)$$

where N denotes the number of samples and $\bar{a}$ represents the mean. From this a covariance matrix is generated:

$$C = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_z^2 \end{bmatrix}$$

The transformation matrix **D** is then a product of: $D = AV$ where $D = (d_1[n], d_2[n], d_3[n])$, $A = (a_1[n], a_2[n], a_3[n])$, and V is the matrix of eigenvectors.

Once the transformation matrix is completed in the preprocessing step, the authors then extract the following features from a 5 second window: mean, standard deviation, FFT spectral energy, frequency domain entropy, and the log of FFT [1]. Additionally, the autocorrelation function of all accelerometer signals was computed. The autocorrelation function is computed by:

$$r_i[t] = \frac{1}{N-1}\sum_{n=1}^{N}\frac{(d_i[n]-\bar{d}_i)(d_i[n+t]-\bar{d}_i)}{\sigma_i^2}$$

The mean, $\bar{d}_i$, of the orthogonal references axises was found to be of little use. However, the mean of $\bar{r}_i$ proved more useful. The variance for both σ^2_{di} and σ^2_{ri} were computed. Most of the activities recognized by this research are periodic, thus there is a need to identify the period. Period identification is a three step process that involves finding the samples that are local maxima, compute the time difference between successive maxima, and estimate the period of the signal as the median inter-maxima delay. The inverse of the median inter-maxima delay is the frequency. When attempting to find a linear equation for the correlation coefficient function, the following equations were used:

$$R^2 = 1 - \frac{S_e}{S_t}$$

where

$$S_t = \sum_{n=1}^{N} (r[n] - \bar{r})^2 \text{ and } S_e = \sum_{n=1}^{N} (r[n] = f[n])^2$$

Having tested the activity recognition algorithms with the above model and equations, Anjum and Ilyas found the autocorrelation functions provided more accurate recognition results than transformed signals, which is computationally beneficial as the autocorrelation functions are cheaper than transformation. The features found to be most useful were the mean as noted above, variance as noted above, standard deviation, R squared, and the period. The majority of this research review focused on the features extracted from the accelerometer data because Anjum and Ilyas found the gyroscope to be of no value in their recognition algorithms. As a note, Anjum and Ilyas found that ascending stairs was the most difficult activity to recognize accurately, the inclusion of a barometer in more cell phone models should thus become a sensor of value when attempting to differentiate between ascending, descending, and non-inclined walking.

2.7 Perfecting Gravity Recognition

In another activity recognition paper, Wang, Chen, and Ma [44] compared between using acceleration synthesization as done by Reddy et al. in [35]:

$$A_{mag} = \sqrt{(A_x)^2 + (A_y)^2 + (A_z)^2}$$

and acceleration decomposition as advocated by [32] :

$$p = \frac{d \cdot v}{v \cdot v}$$

Wang et al. focus on accelerometer data for activity recognition is based on the previously stated premise that the accelerometer is signal independent (unlike GPS), has low energy consumption, has instant startup, and as such is a wholly contained sensor with no external requirements. Wang et al. extracted the following features: mean, standard deviation, mean crossing rate, third quartile, sum and standard deviation of frequency components between 0Hz and 2Hz, ratio of frequency components between 0Hz and 2Hz to all frequency components, sum and standard deviation of frequency components between 2Hz and 4Hz, ratio of frequency components between 2Hz and 4Hz to all frequency components, and spectrum peak position for a total of 11 features. These 11 features were used for the synthesized accelerometer data. For the decomposed data, the 11 features were applied to both the vertical and horizontal axises with an additional feature added for the correlation coefficient between the two series, leading to a total of 23 features. After the experiments and analysis were performed, Wang et al. found that the SD features produced more accurate results than the decomposed. Wang et al. surmised that if the window length is not long enough or the estimate for gravity is not accurate, the decomposition technique will yield features not as viable for accurate activity recognition as the synthesized method.

Using a DT, Wang et al. found that using the decomposition technique yielded an accuracy of 60.71% and the synthesized technique yielded and accuracy of 61.42%. Pairing down the feature set through the Waikato Environment for Knowledge Analysis (WEKA) machine learning library of algorithms changed the decomposition and synthesized results to 60.43% and 70.73% respectively.

Activity recognition via attached sensors as a science has undergone continual refinement, with the placement of powerful and versatile sensors in cell phones the pace of refinement is rapid. In an extension of the previous work [44], the Hemminki, Nurmi, and Tarkoma work to improve the gravity component found to effect the accuracy of accelerometer decomposition [18]. Noted Hemminki's et al. discussion of previous work is that accelerometer synthesization is accurate for pedestrian activity detection, the technique is less accurate for detecting motorized activity. According to the research, only accurate decomposition offers the fine-grain features necessary to observe the acceleration and deceleration patterns of various motorized transportation mechanisms. As such, the Hemminki et al. worked to improve the computation of the gravity component. The goal is similar to earlier work [19] where the horizontal component was computed. Knowing accurate vertical and horizontal axises allows for more accurate identification of acceleration and deceleration periods; introduced are the concept of *peak* features to characterize acceleration and deceleration pattens associated with different motorized modalities.

As synthesization works well with accelerometer data, synthesization should work equally as well for magnetometer and gyroscopic data. By using synthesization in the research into entity recognition, the need to decompose the various sensor streams is eliminated, resulting in classifiable data with less data processing. In a situation where an entity effects the magnetic field detectable by the smart phone sensors, if one axis on the sensor detects field, it is likely the other two axises would detect some magnetic change as well. By synthesizing the data each of the axises are combined into a single output thus the need to assign attributes to each axis is negated.

As accelerometers are the principle sensor utilized when discussing physical activity recognition, there has not been much mention of fusing other sensor data into the algorithms on more than a minimal basis, when doing so added to the fine-grain classification efforts. The work of Barthold, Subbu, and Dantu in Evaluation of Gyroscope-embedded Mobile Phones explores the exploitation of gyroscope data to determine device orientation [2]. Barthold et al. believe that the accelerations experienced by the phone limit the usefulness of accelerometer data in determining device orientation. While much of the previously discussed work has been about how to make activity recognition orientation and placement agnostic, this work is oriented more towards understanding the precise orientation of the device. Once a determination has been made on the precise device orientation, the inertia experienced by the gyroscope can then be used to infer direction changes. Typically the accelerometer and magnetometer sensors are used as multi-modal sensors to determine device orientation. Using the gyroscope to infer direction changes can be useful in environments such as indoor and/or urban environments, environments that will compromise the ability of the magnetometer to determine device orientation. The major complication when using gyroscope data is that gyroscopes tend to exhibit drift areas over time, as such the drift errors result in a decrease (or increase) in a final result in a given time window, thus a process needs to be put in place to account for the drift. If the drift error can be overcome and neutralized, the benefits to adding gyroscope data to device orientation determination is that the gyroscope is immune to external accelerations and magnetic interference, thus algorithms will be able to determine orientation even in magnetically interfered areas while the phone is accelerating.

The use of gyroscopes to determine smart phone orientation is an example of multimodal orientation detection, as Barthold et al. proved that both the accelerometer and gyroscope are capable of determining a smart phone's orientation with varying degrees of accuracy. Additionally, in perfecting the gyroscope orientation technique the magnetometer was used to obtain pertinent magnetic readings, demonstrating the relevance to multimodal sensor fusion in smart phone when it comes to detecting device orientation. The multimodal sensor fusion used in entity recognition goes past smart phone orientation to entities complete external from the smart phone.

2.8 Multimodal Success

A more recent addition to the concept of sensor fusion is the CMOS sensor based camera present in cell phones. In the paper, Using CMOS Sensors for Gamma Detection and Classification, Cogliati, Derr, and Wharton explore using a standard cell phone to detect gamma radiation [6]. The CMOS facilitates the detection of ionized electrons; when ionized electrons are emitted by a gamma emitting object and make contact with a cell phone's CMOS sensor, the sensor is capable of registering this particle strike. The detection is based on the principles of scattering and ionizing radiation, as well as the different energy levels associated with various types of radiation. Using CMOS sensors to detect electrons has a large noise correction requirement, as a CMOS sensor will detect electrons due to leaky circuits in the phone. Heat will increase the amount of electrons emitted from leaky circuits. The detection of leaky circuits is a fairly static feature that will be visible in subsequent images, as such filtering will remove similar detections. Gamma rays on the other hand continually produce a stream of ionized electrons, due to the nature of scattering the CMOS detections will strike different locations on the sensor and will not emulate leaky circuits. In addition to compensating for thermal noise, Cogliati et al. found the need to compensate for defective pixels. As a CMOS captures images in three colors, red, blue, and green, it is rare that at a given pixel location all three receptors are bad. As such, the preprocessing algorithms verifies each component of each pixel individually to determine whether it is functioning. Once leaks and bad pixels have been identified, a number of noise removal techniques were assessed to account for signals that didn't correspond to identified leaks and defective pixel components but still may be erroneous. Cogliati et al. used median value noise reduction, statistical methods using the standard deviation and mean (background = $max\{y \in x \| y < \bar{x} + 2\sigma_x\}$), kurtosis, and the high-delta method. The high-delta method takes the max value and second highest value seen in a set of images and finds the difference between the two values, thus reducing both thermal and defective pixel noise. Cogliati et al. found that using the cell phones CMOS sensor and phone based data processing of images, the cell phone is capable of functioning as a low-sensitivity dose rate meter with limited spectrum information. While not particularly active compared to dedicated meters, the ubiquity of cell phones makes it useful when other tools are not available. As a sensor fusion example, an application (GammaPix) have been designed where the GPS, accelerometer, and CMOS data streams have been co-utilized to locate airports (GPS), detect takeoff and landing (via accelerometer data), and monitor high-atmosphere radiation exposure (via CMOS). Entity recognition seeks to expand on the concepts utilized in the gamma radiation detection methods of GammaPix by sensing not just environmental phenomena but also the entities producing the phenomena.

2.9 Smart Phone Flocks

An example of monitoring the movements of groups of people via sensor fusion can be found in Detecting Pedestrian Flocks by Fusion of Multi-Modal Sensors in Mobile Phones [24]. While much of the prior discussion has been focused on the concept of recognizing activities performed by individuals, this work focuses on the joint identification of the indoor movement of multiple people forming a flock. A flock can be thought of as a group of persons moving in the same direction for some duration, or more formerly as the existence of a moving cluster with regards to the ground truth location data. It can be thought of algebraically as a pedestrian flock $\mathbf{F}$ is a moving cluster that exists for the duration $t \geq \tau$ and consists of more than $n \geq \nu$ people where τ and ν are application specific. Kjrgaard, Wirz, Roggen, and Tröster found that combining sensors in a multi-modal fashion improved the accuracy over unimodal approaches. The multi-modal approach allows for robustness when a single category of sensor may fail; detection accuracy improves in the multi-modal approach and energy savings may be achieved through specific combinations of sensors for detecting flocks in specific environments. One scenario where the detection of a pedestrian flock is desirable is to aid emergency personnel during evacuation processes. In addition to aiding emergency personnel, it would be beneficial to target the flock with location and movement appropriate messaging.

Identifying a pedestrian flock is performed via a cluster-based weighted majority voting system. A weighted majority voting is performed that outputs a set of clusters that the majority of features agree on. As the flocks are clusters of people where a majority stay together over time, temporal clustering is performed to combine highly similar clusters that exist for several successive time windows into flocks. This process allows Kjrgaard et al. to output devices grouped into flocks, and thereby people as well [24]. Features will be generated from the data output by the accelerometer, magnetometer, and GPS sensors, as well as the WiFi radio. Accelerometer features are used to correlate movement and acceleration variance similarity between potential flock members. Magnetometer features are used to correlate turn and relative heading changes, the similarity of the changes are compared between potential flock members. GPS data is used to examine proximity, speed, and heading differences from location-fingerprinting when available. The WiFi radio signal feature is observed to determine similarity in signal strength. Performing pair-wise correlation on the above features will result in a $n \times n$ matrix for each feature. Then, weighting the features abstracted from the four sensors helps to identify cell phones that belong to the same pedestrian flock. This is performed with both spatial and temporal clustering.

Using the accelerometer data to detect pedestrian flocks, Kjrgaard et al. utilize Overlap in Movement Behavior (OMB) and Windowed Cross-Correlation of Acceleration (WCCA) algorithms [24]. OMB is used correlate cell phones that exhibit similar activity; activity recognition is performed at a rather course-grained level in their research, identifying stationary vs moving activities. After computing two lists of moving and stationary entities, M_a and M_b, respectively, the following OMB similarity feature computation is performed over a specified time window:

$$S_{a,b} = \frac{\sum_{t=t_0-T}^{t_0} f(M_{a_t}, M_{b_t})}{n}$$

Using the WCCA method, Kjrgaard et al. are able to analyze acceleration signals to determine whether two signals are the result of similar movement behavior. The WCCA method considers variance of the signal magnitude to mask variations in device orientation and small differences in movement trajectories. This method allows flexibility in cross-correlation as members of a flock are not constrained to walk in step, thus the similarity of a pair of devices will be computed from measurement streams of acceleration magnitude,;two lists of acceleration magnitudes are computed, one for each device being compared. Since behavior changes can be shifted in time between flock members, the maximum cross correlation is computed with a lag between minus one and plus one second. As such the WCCA is computed by:

$$S_{a,b} = max(corr(V_a, V_b, \ell), \ell \in [-1, 1])$$

Magnetometer data features are used to determine whether individuals walk together as measured by the phone's magnetic orientation; Kjrgaard et al. use Windowed Cross-Correlation in Relative Heading (WCCH) changes and Time Since Last Turn (TSLT) algorithms [24]. Similar to the WCCA used for accelerometer features, each device's heading is computed and compared to another. As such two lists of heading deviations, H_a and H_b are cross correlated with:

$$S_{a,b} = max(corr(H_a, H_b, \ell), \ell \in [-1, 1])$$

The TSLT algorithm first detects turn then computes the duration of time between turns to determine whether similarity exists. Turns are computed by comparing the mean compass orientation measurements for devices:

$$y = mean(w_1(C_a)) - mean(w_2(C_a))$$

against the standard deviation of the compass orientation measurements for the devices:

$$y \geq \frac{StdDev(w_1(C_a)) + StdDev(w_2(C_a))}{2} + G$$

where G is a guard factor. From this information, Kjrgaard et al. compute a list, K, for each device which has a zero for when a turn is detected and else the previous value:

$$S_{a,b} = \sum_{t=t_0-T}^{t_0} \|K_{a_t} - K_{b_t}\|$$

WiFi features are analyzed to determine spatial features where WiFi positions are applied to a predefined map of signal strength measurements and to determine signal strength features where flocks are detected. The spatial features model the similarity between two mobile devices as the shortest walking distance between their position via the predefined map; devices that have larger walking distances will be less likely to be clustered. Signal features are computed for devices based off their signal strength vectors and compared to other devices to derive their Euclidean distance. Additional signal features are SpatialSpeed and SpatialHeading that are computed as the minimum sum of differences in speed and heading within a window of time. WiFi features help to identify an individual's location with location-fingerprinting and signal strength (when available), and the SpatialSpeed and SpatialHeading are able to be cross-correlated to determine device movement similarity, thus the WiFi features aid in providing finer-grain detection of pedestrian flocks.

Having chosen the feature correlation algorithms, Kjrgaard et al. explore different clustering techniques to determine whether pedestrian clusters exist [24]. Using the geometric features found in the WiFi spatial and signal features, Kjrgaard et al. use a hierarchical clustering algorithm. Using the non-geometric features (e.g., the rest of the features), Kjrgaard et al. use a density clustering algorithm. The clustering is run against the similarity matrixes described previously for each feature. Once the clustering has been performed, the clusters are fused to improve the overall quality, this is done by weighted majority voting to combine clusters identified in the different feature spaces. The weighting is based on the quality of the selected features. In order for devices to become members of the same flock, the devices must exhibit feature sets and quality, and they are required to have membership in successive time stamps to join a flock. Flock recognition was most accurate when using OMB, TSLT, spatial, and signal features, thus a fusing of accelerometer, magnetometer, and WiFi produced the most accurate results. When wireless access points (WAPs) were not available, or the location-fingerprinting was not achievable, OMB and TSLT performed best. GPS did not prove worthwhile to fuse due to the indoor nature of the research performed.

2.10 Natural Event Entity Recognition

Using the sensors within a cell phone for detections beyond the human activity realm is an area of research ripe for study. Dr. Faulkner et al. have developed a process to utilize cell phone sensors to monitor for external environmental events, namely earthquakes [12] [11]. Faulkner et al. research in [12] lays the foundation for detecting events that are difficult to model and characterize a priori with heterogenous, community-operated sensors. As envisioned, each sensor detects unusual observations and will notify a *fusion* center of such observations. Determining what an unusual observation threshold is typically relies on conditional probabilities such as:

$$\frac{\mathbb{P}[X_{s,t}|E_t = 1]}{\mathbb{P}[X_{s,t}|E_t = 0]} \geq \tau$$

However, an event such as an earthquake, due to its' rarity, does not have sufficient data to obtain good probability models. In addition, since the composition and placements of sensors is heterogenous, each will record varying environmental factors. Lastly, while it is plausible that much of the higher math could be performed at the fusion center to determine whether an event has taken place, bandwidth limitations and resource availability necessitate developing a more reliable method for event detection at the cell phone level. Faulkner et al. developed a *pick* method whereby the transmission of false-positives to a fusion center could be mitigated. Using a likelihood specific to each sensor, a variation of the above probability threshold will determine whether a signal is sufficiently different from normal data so that the probability of an event taken place is significant:

$$\frac{\mathbb{P}[x|E_t = 1]}{\mathbb{P}[x|E_t = 0]} > \frac{\mathbb{P}[x'|E_t = 1]}{\mathbb{P}[x'|E_t = 0]}$$

thus the less probable x is under normal data, the larger the likelihood ratio gets in favor of the anomaly. In order to get this equation to work as desired, Faulkner et al. have to establish the parameters for a sensor to estimate the distribution in an online manner, establish a sensor specific threshold for anomaly recognition, and then develop the true positive, false positive, and appropriate anomaly threshold rate for the fusion center.

In order to establish an online density estimation for each sensor, Faulkner et al. develop a methodology to estimate the distribution of normal observations over time $\hat{L}_0(X_{s,t}) = \hat{\mathbb{P}}[X_{s,t}|E_t = 0]$ for non-events. This is done by using a parametric approach:

$$\mathbb{P}[X_{s,t}|E_t0] = \phi(X_{s,t}, \theta)$$

This model improves when the time span of sensing increases and thus the availability of training data increases. The fusion center can send back updated θ to each device in order to improve their detection algorithms. In order to set the online threshold estimation for a specific sensor so that the per-sensor false positive rate can be controlled, an appropriate τ_s must be chosen. Using the ϵ-approximation to limit the search space

$$\frac{|r' - r|}{N} \leq \epsilon$$

then assuming that τ_s is obtained through a percentile estimation for p_o, τ_s can be found by

$$\hat{p}_0 = \mathbb{P}[\hat{L}_0(x_{s,t} < \tau_s]$$

The two above probability functions complete the variables necessary for earthquake detection on a cell phone. Without getting into the algorithmic process present at the fusion center, the method by which the network identifies earthquakes will be discussed [12]. After each sensor has *learned* the decision rules that allow for the control of system-level false positive rates, each sensor decides on its' own whether it believes an event has taken place. When a sensor believes an event has taken place, it sends a *pick* message to the sensor fusion center. The fusion center will then decide whether an event has occurred by comparing the number of sensors reporting **1** for an event versus **0** for a non-event.

The task of detecting the earthquake is left to the accelerometer; accelerometer and location data are fused to report on acceleration values where location is determinable [12]. The earliest iteration of the work required the cell phone to be plugged in and laid down order to sense events, this allowed for plenty of computational resources as well as a stable setting where human interaction would have minimal impact on sensing. Experiments comparing earthquake acceleration values against the standard deviation of resting sensors have revealed that an earthquake that registers 4.0 on the richter scale would be the minimum detectable by a cell phone sensor. As in previous research, signal rotation is necessary to determine the estimated gravity components in the negative z-axis. The *picking* algorithm could then be utilized to analyze live data to determine whether it is anomalous. The pick data would be sent to the cloud fusion center (CFC). Using received *picks* and a geographic hashing, the CFC would send *heartbeat* messages to nearby phones to determine whether they are active or not. The geographic hashing would ascribe integer hashing to a grid of latitude/longitude headings, and the grid cell size would be determined by the propagation rate of seismic waves and the feature calculation window dictated by the extraction algorithm. In addition, the geographic cells that received *picks* would be put into time windowed buckets at the CFC for processing. With the received *picks*, the location of the picks on the grid, and an arrival time captured, the CFC works to probabilistically determine whether an event has taken place.

2.11 Identifing Clusters of Importance

In a work from 2004, researchers investigate the use of location aware cell phones and interactive clustering in the development of a personal gazetteer to identify and locate important destinations [46]. Zhou, Frankowski, Ludford, Shekhar, and Terveen identify an individual's most important places (e.g, home, work, grocery store, etc.). Zhou et al. developed an application to capture user's locations throughout the day; from this set of location data, an algorithm determines which data points represent a cluster and thus indicate proximity to an important place. Non-deterministic approaches such as K-Means clustering and deterministic approaches such as density based clustering were both considered. A density-based deterministic algorithm was chosen as it allows cluster of arbitrary size, robustly ignores outliers, noise, and unusual points, and provided deterministic results.

$$N(p) = \{q \in S | dist(p, q) \leq Eps\}$$

The density-based clustering algorithm uses temporal pre-processing techniques to reduce the number of uninteresting places that are discovered; as such locations with speeds greater than zero and locations of close proximity to another reported location are discarded, greatly reducing the amount of data. Additionally, the preprocessing step would aid in the removal of frequent (and similar) stop locations that may exhibit inconsistency in zero speed readings (and location parameters) such as traffic lights. Then the density-based algorithm can comb through the spatiotemporal history using the time-stamped location data to *discover* the personal gazetteer. When combing through the data, significant events can be detected by the loss (or gaining) of GPS signals, as this indicates the entering (or departing) a building or similar structure. This GPS signal change makes the use of a clustering approach unnecessary for the detections of certain places, but to detect locations such as parks, stadiums, or sidewalk cafe where a GPS signal is constant, the density-based algorithm proves necessary. Applications of this research extend beyond the concept of personal gazetteers and into the realm of partner matching for car pooling/transportation needs.

As data clustering has presented itself as a necessary technique for the recognition of events, locations, and entities, a review of techniques is presented in [14]. Clustering (or grouping) of common elements is accomplished in either an exploratory or confirmatory manner based on either a natural grouping (identifiable through analysis) or goodness-of-fit (as a model postulates). Clustering is accomplished through a five step process involving pattern representation, definition of a pattern proximity, the clustering or grouping, data abstraction, and the assessment of the output. Pattern representation refers to the number of patterns identifiable by the clustering algorithm.

Pattern Representation → Pattern Proximity → Clustering → Abstraction →

→ Output Assessment

A set of features is presented to the algorithm to utilize in the identification of patterns. The selection of the features is the process of identifying the most effective subset of features to utilize in clustering. Feature extraction is the use of one or more transformations of the input features to produce new features. The use of feature selection and/or feature extraction is often the crux of most recognition research; considerable effort is made to identify the features sets that produce the best results. Patterns can be based on either quantitative or qualitative features. Quantitative features are typically continuous values, discrete values, or interval values. Qualitative features are nominal or unordered and ordinal values.

Pattern Representation → **Pattern Proximity** → Clustering → Abstraction →

→ Output Assessment

Pattern proximity is measured by a distance function defined on pairs of patterns. Euclidean distance is simply one variety of distance measure used to determine how similar two patterns are to one another. Patterns that are closer together share a higher likelihood of sharing a classification as compared to those that are further apart. Euclidean distance can be found by:

$$(\sum_{k=1}^{d} |X_{i,k} - x_{j,k}|^p)^{1/p}$$

of which there are a number of different derivatives based on the features being compared.

Pattern Representation → Pattern Proximity → **Clustering** → Abstraction →

→ Output Assessment

Clustering or grouping can be performed in a number of ways. In hard clusters, clusters are separated by a partition whereby the data is grouped according to some common property. In fuzzy clustering, clusters may vary and depend on varying association with a set of patterns, as clusters may share properties with multiple patterns. These clustering techniques can be further categorized as hierarchical or partitional. In hierarchical techniques, algorithms produce a nested series of partitions based on merging or splitting criterion. Partitional clustering identify the partition that optimizes a particular criteria (usually at a local level). Additionally, probabilistic and graph-theoretic clustering techniques are described by P.J. Flynn in section 5 of his work[14].

Pattern Representation → Pattern Proximity → Clustering → **Abstraction** →

→ Output Assessment

Data abstraction is the process of abstracting a simple and compact representation of a data set. Abstraction is performed in order to achieve efficient machine based processing for output assessment or by representing the data in an easy to comprehend manner for human-oriented review.

Pattern Representation → Pattern Proximity → Clustering → Abstraction →

→ **Output Assessment**

Output assessment is the processing of confirming cluster validity. If the output of a clustering algorithm is unusable, one of the four prior steps needs to be reimplemented.

Data clustering techniques can be further broken down into the following taxonomies [14]: agglomerative vs. divisive, monothetic vs. polythetic, hard vs. fuzzy, deterministic vs. stochastic, and incremental vs. non-incremental. An agglomerative approach begins with each pattern in a distinct cluster and successively merges clusters together until a stopping criterion has been satisfied. A divisive approach begins with all patterns in a single cluster and performs splitting until a stopping criterion has been reached. In a monothetic approach, the algorithm considers features sequentially to divide the given collection of patterns by distance. A polythetic approach is where all the features available to an algorithm enter into the computation of distance between patterns. The polythetic approach is used far more often than monothetic since the overall distance measured in monothetic will vary according to the order of feature comparison. Hard and fuzzy techniques were described in the previous paragraph and are related to the degree of inclusivity a pattern has with a classification. Deterministic algorithms use traditional algorithms whereas stochastic algorithms resort to more randomized algorithm such as genetic or evolutionary algorithms. An incremental approach evaluates patterns one at a time and functions best for small data sets with a minimal number of classifications, thus a method that employees incremental algorithms should work to minimize the number of scans through a pattern set, reduce the number of patterns examined, or reduce the size of the data structure used. A non-incremental approach is utilized when constraints on execution time or memory space affect the architecture of the algorithm. Choosing the right approach to clustering is an important step in recognition activities and is guided by the sensors being used and the features abstracted from the data generated by the sensors.

2.12 Multimodal Activity Recognition

In the research paper Comprehensive Context Recognizer Based on Multimodal Sensors in a Smart-Phone, Han, Vinh, Y. Lee, and S. Lee seek to fuse the optimal combinations of sensors together in order to determine the user's context (activity) [17]. Using multiple sensors, namely accelerometer, audio (microphone), and signal (GPS, WiFi), Han et al. work to increase both the number of activities recognized, but also the ability to recognize multiple activities, such as the ability to recognize someone using the cell phone to make a call while walking. This ability to recognize context within context has benefits for resource preservation. As an example, the system utilized the accelerometer to detect transition points from pedestrian activities to transportation activities, and vice versa. When the accelerometer detects transportation, the WiFi receiver may be asked to identify private WiFi connections which will be far more common on a bus than a subway. For instance where the accelerometer is not able to provide the fine-grained detail needed in this study, the audio classifier would be enabled to further classify a transportation activity. Additionally, by identifying the feature sets best able to identify activities, sensors that generate data for unused feature sets can be disabled till the a context change is detected.

Utilizing multimodal sensors requires a balance in classifier selection. In some instances where the data streams and features are similar, such as accelerometer, gyroscope, and magnetometer data, the same classifier can be chosen. In cases where dissimilar data output sensors are chosen, such as accelerometer and microphone, multiple classifiers will be required. In Han's et al. research into multimodal sensors, they chose a Gaussian Mixture Model (GMM) and a Hidden Markov Model (HMM) for the accelerometer and audio classifiers, respectively [17]. The GMM allows for the use of multiple dimensions of features where there may be multiple distributions of the data represented. The HMM was chosen for the audio classifier as there are only two audio signature being detected and distinguished between, the bus and subway. Unlike the super fine-grained accelerometer approach to recognizing the differences between acceleration and deceleration patterns in buses and subways demonstrated in [18], Han et al. utilize a more coarse classifier that activates the audio classifier when more fine-grained detail is required. This difference in approaches demonstrates the flexibility present in the suite of sensors available in cell phones. The features extracted for activity recognition vary among the research. The best features are selected from the following features: standard deviation, mean crossing rate, Pearson correlation coefficients, frequency domain features, and linear predictive coding features to name a few. Due to the 'curse of dimensionality', using all available features would not necessarily result in a more accurate recognition of activity, as such it is prudent to select the *best* features. Han et al. have built an algorithm that seeks to select features based on two qualities: the first being the relevancy of the feature (or the classification power) and the second being the redundancy of the feature (or the similarity of two features). Once the relevance and redundancy have been calculated for each feature, a *greedy forwarding* search technique is applied to selectively extend the feature set for inclusion into the classifier suitable for that sensor's data.

In the research Preprocessing Techniques for Context Recognition from Accelerometer Data, Figo, Diniz, Ferreira, and Cardoso provide additional scenarios where the use of activity (context) recognition proves useful [13]. Additionally an overview of numerous features is discussed in detail for the time, frequency, and discrete representation domains. An addition to the concept of recognition activity, Figo et al. advocate that by analyzing an individual's activities over the course of days, weeks, or months, a more interactive experience can be offered to users. A couple of instances are the ability to aid the elderly and the ability to offer value-added information. In the case of elderly aid, if an awareness of a user's activity could correlate an abrupt change as a potential red-flag, such as an elderly individual taking a fall, it is conceivable that an emergency service could more easily and accurately be made aware of the situation. As a value-added situation, consider the case of an activity recognizer that *knows* an individual runs at a certain time of day or on a particular route. If the activity recognizer can correlate this information with a weather forecast or traffic construction, it is conceivable that the user could receive suggestions to alter their time or route.

2.13 Attribute Selection

In order to offer value-added services to an user, it is necessary to possess the ability to acquire, manage, process, and obtain useful information from the raw sensor data. From this sensor data, devices must be able to accurately discover the characteristics or features of the signal coming from the sensor. Figo et al. discuss the layered architecture responsible for this task:

Sensor Data $\rightarrow$ Preprocessing $\rightarrow$ Sensor State $\rightarrow$ Classification $\rightarrow$ User Context $\rightarrow$

$\rightarrow$ Applications

Within the **preprocessing** layer there are effectively two layers: the base layer that determines whether there is a specific short-term context or state, and the base-level classifier to determine the type of activity being performed. The preprocessing is split as it is easier to identify a short-term context such as the absence of light or the presence of a quick movement (like a fall) and it is more computational intensive to accurately identify and classify a specific type of exercise. In processing the signals for features, Figo et al. explore features related to the time, frequency, and what they call the discrete representation domains [13].

Time domain features are those that are derived via simple mathematical and statical metrics from the raw sensor data. These techniques compute features from the sensor data according to some determined time window. The most common features available in the time domain are the mean, median, variance, standard deviation, min, max, range, RMS, correlation, cross-correlation, and the integration features. The **mean** is calculated over some window and is typically used to determine a user's posture and whether an activity type is static or dynamic. The mean is also used as a preprocessing component as knowing the mean aids in the removal of random spikes and noise, smoothing the overall dataset. The **median** is utilized to replace missing values. The **variance** is the average of the squared differences from the mean and is utilized where a threshold is required for classification. The **standard deviation** is the square root of the variance and represents both the variability of the dataset and a probability distribution. The standard deviation is an indication of the stability of a signal, however, it becomes less useful if spurious values are included. Taken together the variance and standard deviation are often used as a signal feature to infer user movement. The **range** can be use with other indicators to distinguish between similar activities, such as running and walking, that will differ in amplitude. The **RMS** is used to classify wavelet results such as those identifiable in walking and biking, additionally the RMS has proven useful as an input for neural networks. The **integration** metric measures the signal area under the curve to obtain speed, distance, and in conjunction with the RMS signal, the ability to calculate the angular velocity from the gyroscope. The **signal correlation** is used to measure the strength and direction of a linear relationship between two signals. The correlation is useful for differentiating between two activities that involve translation into a single dimension. The degree of correlation requires calculating the correlation coefficient and is used to determine which classifiers are the best for recognizing activities. **Cross-correlation** is the measure of the similarity between two waveforms and is used to search for a known pattern in a long signal.

Additional time domain features are the differences, angular velocity, zero-crossings, Signal Magnitude Area (SMA), Signal Vector Magnitude (SVM), and the Differential Signal Vector Magnitude (DSVM). Sample **differences** allow for the basic comparison between the intensity of user activity when arranged pairwise. **Zero-crossings** are the points where a signal passes through a specific value corresponding to half of the signal range and are used for recognition of step movements and the detection of appropriate timing for the application of other techniques. Zero-crossings are used in conjunction with HMM to detect complex human gestures. Angle and **angular velocity** are used for detection of user orientation and has proven useful for fall detection as well as location detection through gyroscopic means [13]. **SMA** are used to compute the energy expenditure during periods of activity. Additionally, SMA can be used to distinguish between resting and user activity. SMA is often used in conjunction with SVM to identify possible falls and classify behavior patterns and with DVSM for dynamic activity recognition using thresholds and single metrics.

Frequency domain features are used to capture the repetitive nature of a sensor signal. The repetition often correlates to the periodic nature of a specific activity. Commonly used frequency domain features are generated from the Fast Fourier Transform (FFT) and the Fast Time Fourier Transform (FTFT). Frequency domain features are the DC component, spectral energy, information entropy, spectral analysis of coefficients, wavelet analysis, and symbolic string domain analysis. Using FFT it is possible to derive frequency domain features similar to those obtained in the time domain, such as averages and dominant frequency components. Using the FFT, the **DC** component is generated and co-utilized with other signal characteristics to determine activity. **Spectral energy** is the energy of a signal and is used during single axis accelerometer activity recognition, and during operations to determine context through audio recording. **Information entropy** helps to differentiate between signals that have similar energy values but correspond to different activity patterns. Together with the mean, energy, and correlation, information entropy has been used to classify activities that contain similar energy levels. **Spectral analysis** of specific coefficients has been used to aid in activity recognition. Using the coefficient of magnitude and frequency peaks within specified frequency ranges, the determination of step rates has been accomplished via spectral analysis. **Wavelet analysis** can be used to examine the time-frequency characteristics of a signal. Wavelet analysis has been used to differentiate and then classify activities that are similar such as horizontal walking versus stair climbing. Transformation into the **symbolic string domain** is used to map signals to strings for matching purposes; it is used to evaluate string similarity and thus find known patterns. In order to facilitate the recognition and classification of symbolic strings, there are three distance formulas used to compute the distance (or similarity) of strings. Euclidean distances are found via:

$$\sqrt{\sum_{i=1}^{n}(|S_i = t_i|)^2}$$

and are used as a distance between symbols. The Levenshtein edit distance allows for the determination of a signal (as a part of a set of possible signals represented as symbols) to determine which is the closest. The Levenshtein edit distance is found via in dynamic programming:

$$d(i,j) = min\{d(i-1,j) + insert, d(i,j-1) + insert, d(i-1,j-1) + subs(i,j)\}$$

where m and n are the length of two strings and d is a $m \times n$ table which is initialized with the costs of creating the input strings. The last distance formula is the Dynamic Time Warping (DTW) process that is used to measure the similarity between two sequences that may vary in length, can thus correspond to different time basis. This DTW approach involves finding the mapping W, where in some case an element of one string can map to sequence of consecutive elements in another string:

$$min\{\frac{1}{K} \cdot \sum_{k=1}^{K} W_k\}$$

where the cost of the post through the cost matrix is found using dynamic programming.

Figo et al. reviewed the suitability of implementing the above features from a quantitative and qualitative approach [13]. Quantitatively they analyzed the complexity of implementation, computational complexity, memory requirements, and precision. Qualitatively they analyzed the suitability for inclusion on a mobile device (cell phone) based on the results of the quantitative analysis. Experimental analysis was performed to determine the best methods to differentiate between walking, running, and jumping. The highest accuracy of activity recognition was found to be generated via features from the time domain. From the frequency domain, coefficient sum and energy exhibited the absolute highest accuracy, but not high enough considering the complexity of implementation and computational cost. When all was said and done, Figo et al. found that the computational simplicity of time domain features indicated all would be suitable except the correlation and cross-correlation features. Figo et al. found the opposite to be true for the frequency domain features. Due to computational cost, only wavelet analysis and the string domain distance finding metric of euclidean distance proved suitable for mobile devices. Figo et al. have provided a comprehensive analysis of numerous features being utilized in the field of activity recognition.

Research into how to select the **best** set of features in A Novel Feature Selection Method Based on Normalized Mutual Information, indicates validity to using either the max-relevance minimum redundancy approach (mRMR) or the Normalized Mutual Information Feature Selection (NMIFS) algorithms to incorporating the most appropriate set of features in an activity recognition model[43]. As noted previously in slightly different terminology, Vinh, Lee, Park, and DAuriol define the concept of feature extraction as the process of generating new features by projecting the original feature space into a reduced-dimension space. Feature selection is defined as the technique for selecting a subset of relevant features, which contain information helpful to distinguishing one classification from another. Feature selection utilizes the concepts of a wrapper, embedding, and filtering. Wrapper approaches make use of the classification accuracy to evaluate the usefulness of features at each step. Vinh et al. found that the need to repeatedly train wrapper based approaches are computationally expensive and thus impractical to utilize for large datasets. Embedded methods of feature selection use particular classifiers to find feature sets. Embedded methods select features in their training phase, but their ability to use a cost function during the feature selection process makes them faster than a wrapper approach. Filter algorithms utilize simple measurements such as correlation to estimate the goodness of features, as a result, filter methods are fast and effective. Filter algorithms seek to find the subset of features that maximizes the following:

$$P_s = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}}$$

Where S is a subset of k features, $\overline{R_{cf}}$ is the mean feature class correlation ($f \in S$) and $\overline{r_{ff}}$ is the average feature inter-correlation. $\overline{R_{cf}}$ and $\overline{r_{ff}}$ are calculated similarly through:

$$\overline{r_{xy}} = \frac{E[(x - \mu_x)(y - \mu_y)]}{\sigma_x \sigma_y}$$

where μ and σ represent the mean and standard deviation respectively. Vinh et al. note that the filter method is not able to describe non-linear relationships among variables where correlation is difficult to establish. In addition, the computation requires that all of the features be numerical values, thus the desire to normalize the information for comparative computation. In order to allow as wide a set of candidate features to be evaluated using a filter technique such as nRMR or NMIFS, Vinh et al. propose to quantize all data prior to evaluation. The quantization algorithm ensures that N levels of data are quantized for each feature requiring quantization. The proposed methodology uses the normalization of mutual information and the feature independent normalized weights to perform the quantization and would be limited strictly to the selection of features criterion only. The process of quantizing the data for comparison of feature sets is done in as computationally simple as manner as possible to limit the utilization of system resources.

The use of low resource algorithms that use easily computed statistical attributes such as range, standard deviation, variance, skewness, kurtosis, and root-mean-squared have been shown to be quite useful when recognizing and classifying activities [18]. It remains to be seen whether such statistical attributes are equally as effective when classifying a variety of non-activity based entities. Additionally, whether the identification methods can identify not just entity categories but also sub-categories where classification between two of the same entities operating at different modes or frequencies is possible. In a possible complication offered by entity sensing, the appearance of non-periodic entities could pose a problem for statistical entity classification.

While activities such as running, biking, and riding a subway may offer periods in time where the activity appears non-periodic, by and large their sensor output (accelerometer and gyroscope) will exhibit periodic functions. When sensing entities, it is conceivable that entities may display similar characteristics when analyzed statistically. The concept of using a smart phone to scan the undercarriage of vehicles that pass overhead may generate a magnetic signature that when viewed as a series of ridges and troughs may be unique between vehicles, but when statistically analyzed the results could be too similar for accurate identification. As such it is necessary to have additional tools by which to differentiate data, this is where wavelets may offer additional resolution.

2.14 Resource Preservation

In an effort to support continuous sensing, Lu et al. of The Jigsaw Continuous Sensing Engine for Mobile Phone Applications, propose a methodology which strives to balance the resource demands of long-term sensing, inference (recognition), and communications algorithms [28]. Lu's et al. jigsaw algorithm, as proposed, preserves the resilience of the accelerometer data processing regardless of phone platform, placement, or orientation. Jigsaw implements smart admission control and on-demand processing for the microphone and accelerometer data; admission control and on-demand processing allow for adaptive throttling of the depth and sophistication of sensing pipelines when the input data is low quality or uninformative. Adaptive pipeline processing allows for judicious triggering of power hungry pipeline stages when appropriate and takes into account the mobility and behavioral patterns of the user to drive down energy costs. Additionally, their platform implements the concept of robust classifiers explored by [35] that allows for different sensors in different placement positions to accurately recognize activity. Additionally, as noted in previous studies, different sensors have different processing costs associated with their unique sampling rates, features sets, and other performance characteristics, Jigsaw tries to optimally balance the functions responsible for sensing with available computing resources.

The Jigsaw platform was developed to utilize sensor-specific pipelines to process data from specific sensors when performing continuous monitoring; additionally it is optimized and able to run completely on the phone without a server requirement

[28]. The accelerometer pipeline is designed based on the fact that sensor sampling is not resource prohibitive and merely requires a robust set of inferences. As such, accelerometer calibration techniques via a one-time user-transparent process are proposed, classification of activities via sub-classification for independence from sensor placement is discussed, and the filtering of extraneous activities and movement is explained. For accelerometer data, misclassification is most pronounced during periods of activity overlap, such as answering a phone while riding. The features selected for classification would be among the following set: mean, variance, mean crossing rate, spectrum peak, sub-band energy, sub-band energy ration, and spectral energy. Such extraneous activity recognition can be countered by recognizing periods of user interaction (phone calls, texting) and by recognizing transitional states (standing up, act of picking phone up). The sub-classification of recognizable activities is enhanced by the use of orientation independent features as much as possible. In addition to orientation independent features, sub-classification of activities allows for the recognition of activities regardless of body placement of the cell phone.

In the use of a microphone, resource consumption, such as memory, computation, and energy usage, are high. Features computed for audio data are: spectral rolloff, spectral flux, bandwidth, spectral centroid, relative spectral entropy, low energy frame rate, and 13 other coefficient features. The microphone pipeline utilizes the concept of admission control and a duty cycle component to regulate the amount of data that enters its' pipeline. When the microphone has detected a sound (signal) that doesn't change for some window (period of time), the microphone will save resources and not perform redundant classification. Additionally, to save computation resources, the microphone pipeline will short circuit the process for common but distinctive sound classes [28].

A GPS pipeline is optimized to learn user activities to budget energy as judiciously as possible. Energy is preserved by recognizing prior activity trends and working to pre-calculate duration to ensure availability of resources when required. The GPS uses a Markov Decision Process (MDP) to learn an adaptive switching schedule for resource consumption. In addition, through fusion with the low-resource consuming accelerometer, the GPS is able to determine additional opportunities to turn off or on. Lastly, not all applications will require constant sensing, as such the GPS pipeline can be tailored to the context being classified.

An example pipeline would look similar to:

$$\text{Raw Data} \rightarrow \text{Preprocessing} \rightarrow \text{Feature Extraction} \rightarrow \text{Activity Classification} \rightarrow$$

$$\rightarrow \text{Smoothing}$$

Preprocessing would consist of:

$$\text{Framing} \rightarrow \text{Normalization} \rightarrow \text{Admission Control} \rightarrow \text{Projection}$$

Feature extraction consists of the feature vector. Activity classification consists of:

$$\text{Activity Classifier} \rightarrow \text{Output Merging}$$

and the smoothing process consists of a smoothing algorithm that performs a simple moving average on consecutive data points in order to minimize the effect of outliers. There are variations between the accelerometer, microphone, and GPS pipelines; the above model captures the pertinent components of each pipeline [28]. Through the accelerometer, microphone, and GPS, and their associated pipelines, a set of inferences and locations is analyzed for the task of activity recognition. The use of pipelines to save on computational resources, and the addition of sub-classification of activities based on phone position, results in a sensing platform that places no burden on the user in terms of calibration, placement, orientation, or awareness of application activation and deactivation.

Through a review of the relevant literature, the evolution of activity recognition can be seen to have progressed from simple, single sensor techniques that differentiated between a few activities to multi-modal systems that fuse sensor data to detect numerous pedestrian and motorized transportation avenues. Additional research has proposed value-added applications to the concept of activity recognition, offering additional services or information depending on the activity being performed. Research into events completely external to the device and unrelated to users, such as earthquake detection, has revealed the utility of the cell phone to be a sensor of more than just user activity. The recognition of ever more activities and entities would extend the evolution of the capability of cell phones to recognize external events. Growing the ability of the cell phone to recognize more requires balancing the activation of available sensors with resources, selecting the best features for classifying specific problems, and preserving the cell phone's *normal* functions while observing, recognizing, and classifying.

III. Methodology - Recognition and Identification

3.1 Experiment Objectives

In designing the experiment explained within below, the goal is to determine whether the sensors in a cell phone are accurate enough to detect and identify environmental actors external to the cell phone, referenced as entity recognition. Entity recognition will allow the cell phone to peer beyond the scope of identifying and classifying which activity a phone user is partaking. If successful, entity recognition will allow the cell phone with its' environmental sensors and the requisite algorithms to become entity aware. With a large enough set of attributes, entity signatures, and cell phone location awareness, the ability to recognize and classify entities presents researchers and analysts with a wealth of data.

As noted in the earthquake research of Faulkner [12, 11] and the gamma ray detection of Cogliati [6], the concept of using a cell phone's sensors to evaluate the environment a user is in is gaining popularity. Beyond using location data to identify crowds and/or flocks [24], the researchers utilize a multi-modal approach and capture accelerometer and gyroscope data to ascertain the likelihood an earthquake took place. In addition to monitoring CMOS for strikes indicative of gamma ray photon emissions, Cogliati's multi-modal approach utilizes GPS and accelerometer data to determine whether a user is at an airport, and then to identify whether they are taking off or landing.

In determining whether there is value to a multi-modal approach to analyzing the environment for entities, an experiment has been designed to collect and analyze data from several entities, both disparate and similar. The data will be gathered by the

SensorSuite program written in the iOS native language of Swift and designed specifically for the purpose of gathering raw data directly from the sensors in the iPhone 5. Once the data is gathered, it will be analyzed to determine whether an entity detection is possible. If the results indicate it is, it will be further analyzed to determine whether there is value added to a multi-modal approach versus using a single sensor.

The experiment captures the conditions affecting the sensors in the cell phone and reveals details about the environment in regard to magnetic field structure and fluctuations (magnetometer), gravitational changes due to movement affects (accelerometer), and torque and inertial affects (gyroscope). In addition to entities creating conditions that affect a specific sensor in a straightforward manner, such as the magnetic field being detectable by the magnetometer, it may be possible to detect the field (and thus an entity) by forces exerted on the gyroscope. In the same vane, it may be that vibrations detectable by the accelerometer and gyroscope based on minute changes in the cell phones orientation could also affect the readings from the magnetometer, as its' location relative to a specific spot in the magnetic field may shift.

It is not known whether a cell phone's sensors offer the fidelity necessary to 'discover' and 'identify' an entity in the environment. Thus, sensor data from the cell phone sensors will be captured and analyzed to determine whether an entity has had detectable affects on the sensor. Apart from the cell phone, it is known that entities produce environmental effects that are measurable and detectable via legacy devices purpose built to sense a specific effect or entity (i.e. seismometers for earthquakes, gaussmeters for measuring magnetic fields). What requires investigation is whether a multi-modal approach to entity detection can augment or replace legacy devices in detection paradigms.

3.2 Experiment Methodology

The experiment built to determine whether a multi-modal approach to sensing entities is obtainable involves 3 distinct groups of control variables for 2 separate experiments. The first experiment involves capturing data from the fused sensor package from the environmental effects induced by microwave ovens and subwoofers. The second experiment involves recording the sensor signature readable from scanning the environmental attributes produced by the undercarriage of a vehicle passing overhead the recording device. It is believed that the microwave oven, active subwoofer, and the vehicle will each produce a magnetic field detectable by the cell phone; it is unknown what effect these devices may have on the accelerometer and gyroscope. Capturing the raw data from each of the three sensors will allow analysis to determine whether a single sensor stream is acceptable for determining which entity is acting on the sensors or does accurate recognition require multiple sensors to determine the classification of the entity. Which entities produce statistically significant affects on a particular sensor beyond a baseline reading where there is no actors save the planetary and structural effects present in the test environment? If classification to a specific entity isn't possible, is it at least possible to get down to the correct category? In order to verify the ability to classify an entity this experiment acquires the raw environmental attribute readings necessary to determine the level of prediction possible.

3.3 Experiment Boundaries

The sensors in a cell phone are regularly used to determine location, phone orientation, motion during app usage, and the particular activity a user may be engaged in, via a combination of data streams from the gyroscope, accelerometer, data communications chipsets (LTE, WiFi, Bluetooth, etc) and GPS. In addition, the sensors (accelerometer and GPS) have proven useful to detecting the presence and non-presence of earthquakes [12, 11]. In theory, it may be possible to use the sensor data to determine the presence of a large number of number of entities in a cell phone's environment. If the data can be combined and/or analyzed in an effective manner, there are whole classes of legacy detectors that could be augmented and/or replaced.

The experiment, as devised, will measure the gravitational, inertial, and magnetic effects an entity produces in an environment that are measurable by the sensors resident in a cell phone. These measurements will be taken by the sensor within the cell phone and captured via the SensorSuite logging software. The measurements will record the effects being read by the sensor as it relates to the environment attributes produced by an entity. The attributes being read by the sensors magnetism, gravity, and inertial effects, are always present in the environment and as such will return a reading. The entity may alter the environmental attributes, if so the sensors within a cell phone may capture the changes.

Each entity involved in this experiment will be measured individually and all reasonable steps will be taken to ensure there is only one entity present and active during a specified data logging session. This is necessary to build a set of data that will allow for the accurate identification of an entity.

3.4 Experiment Response Variables

The response variables for this experiment (Table 1) are the sensors within a cell phone. Using an iPhone 5 as the sensor package, the experiment will log the sensor output from the cell phone's accelerometer, gyroscope, and magnetometer. Additional data from the phones GPS and microphone will be captured for posterity as well, but will not be analyzed in the data analysis phase of this research. The magnetometer, accelerometer, and gyroscope are each 3-axis measurement devices capable of taking readings in the x, y, and z-axises. The magnetometer measures the magnetic field, the accelerometer measures gravitational data, and the gyroscope measure torque and inertial effects. Each sensor is silicon based and determines the environmental attribute it is responsible for via different means. Knowing the specific values captured and output by these sensors and how that correlates to an entity and its' effect on the environment is not straightforward. For instance, the output from a magnetometer can determine where magnetic north is, but the output of its' sensors is not a 180° or 360° output. As such, additional understanding of the physics behind each sensor is required to fully interpret the data output. However, this is not a requirement when it comes to capturing and analyzing the data for statistical significance between actors.

The response variables will be represented in the units native to that sensor. The magnetometer will capture readings measured in µ-Tesla, which represent the magnetic field experienced by the sensor. The accelerometer will capture readings measured in g which represent the gravitational force being experienced by the sensor. The gyroscope will capture readings measured in I which represent the inertial forces being experiences by the sensor.

Table 1. Response Variables

Outputs	Units	Measured Variable
Magnetometer (3-axis)	micro-Tesla (µT)	Magnetic Field
Accelerometer (3-axis)	Gravitational Units (g)	Gravity
Gyroscope (3-axis)	Inertial Momentum Units (I)	Change in Momentum

3.5 Experiment Control Variables

In order to obtain the required environmental attributes via the response variables, the experiment is setup with a number of control variables and held-constant factors. The control variables to be used in the experiments are the external entities. In this case, the experiments will take measurements of the environmental attributes affected by a 12" subwoofer, microwave ovens, and two automobiles. The experiments will be conducted to determine which environmental attributes are affected by the entities when the entities are in an operational status and the cell phone is capturing the attributes via its' sensors. Non-operational (baseline) status readings will be captured as well to register the structural and geophysical properties present in the test environment.

For experiment 1 detailed in Table 2, the recording device (iPhone 5) will be positioned and oriented in the prescribed manner from each entity. For entities 1 through 6, the device will be positioned one inch from the back of the subwoofer enclosure, opposite of the subwoofer; the device is positioned approximately 8.1 inches from the entities magnet. The recording device will be in a head-to-tail fashion with the face of the device pointed skyward. Additionally, the device will rest on the edge of 6x3.5x7.5 inch block of wood, so that the block of wood is no closer than one inch from the subwoofer enclosure; the wooden block is composed of 4 identically sized

Table 2. Control Variables - Experiment 1

Entity #	ID[g]	Entity	Level	Sub-Level	Sessions
1	d	12" Subwoofer	40Hz	dB level 'A' [a, b]	36
2	f	12" Subwoofer	40Hz	dB level 'B' [a, c]	32
3	e	12" Subwoofer	40Hz	dB level 'C' [a, d]	30
4	g	12" Subwoofer	50Hz	dB level 'A' [a, b]	64
5	i	12" Subwoofer	50Hz	dB level 'B' [a, c]	32
6	h	12" Subwoofer	50Hz	dB level 'C' [a, d]	30
7	c	Microwave Oven	1600 Watt	100% Power [e]	30
8	b	Microwave Oven	1000 Watt	100% Power [f]	30
9	a	Microwave Oven	1000 Watt	50% Power [f]	30
10	j	Baseline	n/a	n/a	40

[a] JL Audio Subwoofer, 12W0v3-4, in a "3/4"-inch MDF enclosure built to manufacturers specifications
[b] dB level on 3 inline-device: computer -12dB, receiver -11dB, subwoofer amplifier +10 gain
[c] dB level on 3 inline-device: computer -24dB, receiver -11dB, subwoofer amplifier +10 gain
[d] dB level on 3 inline-device: computer -24dB, receiver -11dB, subwoofer amplifier +5 gain
[e] General Electric, model JES1142SP1SS
[f] Hamilton Beachm model HB-P100N3oAL-S3
[g] WEKA Confusion Matrix ID

pieces of pine 2x4 that have been glued together. This will point the device at the approximate middle of the subwoofer enclosure. For entities 7 through 9, the device will be positioned 6 inches from the front of the microwave, facing the microwave door; the device is positioned approximately 13 inches from the entities magnetron. The recording device will be in a head-to-tail fashion with the face of the device pointed skyward. Additionally, the device will be laid flat on the surface in front of the microwave, is this case a basement floor. For entity 10 the device will be laid flat in the same location as the recording session for the microwave with no other entities present.

For entities 1 through 9, the recording session will be started with the entity in the inactive position, once the recording session is active, the entity will be activated. When the entity has completed a cycle for its' prescribed activity, the recording session will be complete and terminated. The activity prescribed to entities 1 through 6 is to generate the prescribed tone (Table 2) at the prescribed dB level via Katsura Sharewares AudioTest program (version 2.1.2). The wave type is Sine with a 100%

Table 3. Control Variables - Experiment 2

Entity #	Entity	Level	Sessions
1	Vehicle	Subaru [a]	30
2	Vehicle	Ford [b]	30
3	Baseline	n/a	30

[a] 2013 Subaru Crosstreek XV, Premium trim, CVT transmission
[b] 2013 Ford F-150 XLT, Supercab, 4x4, 145" Wheelbase

pulse width at a sample rate of 44.1k for a duration of 3.0 seconds. The activity prescribed to entities 7 through 9 is to operate at the prescribed power level (Table 2 for time durations split between either 30 and 60 seconds; the device was set to heat a bowl of water. Each of these entities, 1 through 10, was recorded at least 30 times.

For experiment 2 detailed in Table 3, the recording device will be positioned and oriented in the prescribed manner from each entity. For each entity, 1 through 3, the device was placed on the same block of wood used in experiment 1 for the subwoofer entity recordings. With the recording device in place on top of the block of wood, entities 1 and 2 were driven at idle speed (varying low speeds) over the wooden block with recording device atop. The vehicle was driven over the block so that the vehicle passed over in a front-to-back fashion with minimal breaking and so that the midline of the vehicle was the approximate passover point relative to the recording device. In addition, the recording device was placed so that at the beginning of each recording session the top of the device faced the front of the vehicle and at the end of each recording session the bottom of the device faced the rear of the vehicle, the device was laid so the face of the device pointed skyward. For entity 3 in experiment 2, the device was laid flat on the wooden block in a residential drive way made of concrete, the same location of the entity 2 and 3 recording sessions.

3.6 Experiment Factors Held Constant

Factors held-constant are the structural environments the readings take place in; in addition the readings are all gathered at approximately the same time thus limiting the amount of change present in potential atmospheric actors. In each experiment, the recording device will be positioned in approximately the same position and orientation to record the entities; the recording location will be marked out on the floor in masking tape. Sans vibrations that move the recording device during a recording session where the entity is active, the recording device and entity will be kept at the distance indicated in Section 3.5.

In order to minimize noise factors as much as possible, the experiment data recording sessions will each have a period of inactivity captured before and after the entity being put in an active status, thus allowing for the verification of *normal* baseline readings. The structural and mechanical noise will be eliminated as much as possible. The geologic noise will be baselined and should not vary greatly over time. The only unknown and uncontrollable, though always present, will be the amount of noise from fluctuations in the Earth's magnetic field. Taken together, the before and after baselining of a particular test will allow for the identification and reduction of noise effects in the environmental attribute readings. In addition, for a sensor such as the magnetometer, it is possible to expose the sensor to a magnetic field of such strength that the sensor requires a software reset to re-baseline itself and may not produce accurate results after exposure to a magnetic field of sufficient strength. The magnetometer experiences sensor overflow when the sum of absolute values of each axis is $\geq 4912\mu T$ [7].

In order to perform the experiment, certain pieces of equipment will be used to ensure replicability of the testing environment. The foremost piece is the recording device, the iPhone 5, with the required internal sensors (magnetometer, accelerometer, gyroscope, GPS, and microphone), this device will be the same for each recording session and will be verified by the unique user identification code that will output with the raw sensor data. Other pieces of equipment will be blocks and tape to outline the testing locations for placement and replacement of the entities and recording devices if movement should occur before, during, or after a recording session. The distances listed in subsection 3.5 will be verified with a standard tape measure with marking in both metric and imperial standards.

3.7 Experiment Data Collection

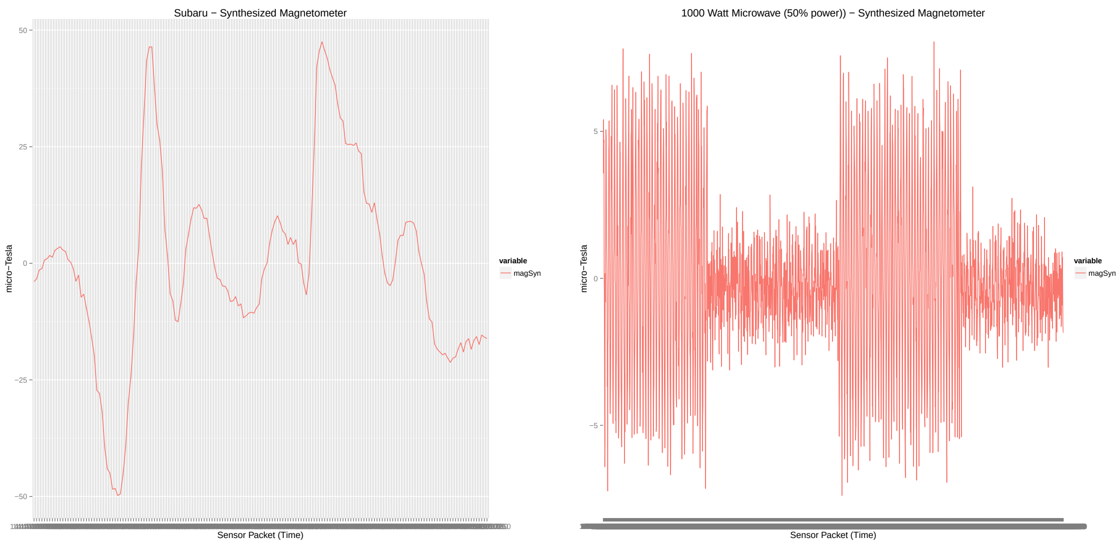
The overall goal of the experiment is to analyze the environmental attributes and how they are effected by specific entities, as such, to capture the environmental attribute output from the cell phone sensors to determine whether the effects are significant enough to measure with the sensors in the iPhone 5. The sensors will read the magnetic, gravitational, and inertial data being output by their respective sensors in the cell phone. These qualities will be recorded and output to a SQLite database at the highest rate possible. The SensorSuite software allows measurements to be captured at rates between 1 and 100Hz a second, the maximum rate possible according to the data sheets available for the sensors within the iPhone 5 [7]. However, the iOS platform limits the sampling rate to approximately 40Hz, presumably for preservation of cell phone resources such as CPU cycles, bus speed, and battery levels.

Each data recording session will capture a single active session for a particular entity, as such, there will be at least 30 recording sessions for each entity. The data samples will be output via a comma separated value (CSV) format for processing in the R statistical computing package. Various screens will be run on the recording sessions to eliminate outliers, trim off the leading and trailing non-active entity sensor data packets, smooth the data when necessary, and to compute the numerous attributes selected for analysis. This process will be repeatable and applied to all data sessions recorded for entity analysis.

3.8 Methodology - Signature Windows

After gathering the data in the experiments listed previously, the sensor data was exported from the iOS SQLITE database via the SensorSuite program designed specifically for the purpose of acquiring data streams from the sensors available on the iPhone 5. The raw data was then imported into R for statistical analysis. Each recording session was date-time stamped by the SensorSuite program for uniqueness and was subsequently broken out separately for analysis. Using the concept of a frame and window approach present in activity recognition [32, 18, 41] to recognize the beginning of a detectable environmental disturbance, the data prior to and after the entity's active period will be trimmed from the data set. The number and length of frames has changed as the field of activity recognition has matured; an accepted standard for activity recognition has settled in the 4-5 frames per over-lapping window, with each frame being comprised of a seconds worth of data samples. Analysis is then performed on each frame to determine which activity is occurring based on the attributes selected. While this approach works well for activity recognition algorithms, it is not well suited for entity recognition.

Using the charts in Figure 1 as an example, we'll examine why frame sizes of one second are not advisable for entity detection. The chart on the left depicts a magnetometer reading for the undercarriage of a 2013 Subaru Crosstrek and took place across a 5.469 second scan (174 sensor plots). The chart on the right depicts a magne-



2013 Subaru Crosstrek Undercarriage (Left) and 1000 Watt Microwave (Right)

Figure 1. Example Magnetometer Signatures

tometer reading from a position 6" inches in front of a 1000 watt microwave operating on the 50% power setting and took place across a 60 second scan (2000 sensor plots).

The x-axis represents the sensor package containing the magnetometer data and is time ordered. Visually it is evident that an analysis at any particular one second frame may not yield the sensor data necessary to determine which entity is acting on the sensors. There are certainly points in each data stream that may be unique to the entity influencing the sensors, but analyzing the entire signature to the the classifier should yield a far more accurate result. It is for these reasons that the frame and window methodology is being used to detect the beginning and end of a signature as opposed to the typical technique of capturing a sample for classification. The algorithm being used to find follows the construct described in the pseudo code depicted in Algorithm 1:

Algorithm 1: Window Algorithm

```
Data: Window of Data
Result: Frame Output
initialization;
Set Window Size, W, Set Number of Frames, numF, Frame Size
sizeF = W/numF, Set n to Data(Length), Set i to Data(Front), Set t
threshold;
j = 1;
while i < (n − W) do
    while j ≤ numF, populate Frame_j do
        Frame_j = (i ∗ j) to ((i ∗ j) + sizeF);
        if if var(Frame_j) > t then
            mark Frame as TRUE
        else
            mark Frame as FALSE
        j++
    if two consecutive frames are TRUE then
        set entity query start to i;
    if two consecutive frames are FALSE then
        set entity query stop to i + W;
    i++
```

For the purposes of identifying a start and stop location to gather data for statistical and wavelet oriented attributes, a window length of twelve ($W = 12$) with three frames was used ($numF = 3$), thus each frame consisted of 4 data points. The threshold was set to 10% higher than the maximum baseline variability reading, as such a threshold of 1.25 μT ($t = 1.25$) was used with magnetometer data to identify the start and stop of an entities potential signature. Similarly, applicable variabilities were used to identify potential start and stop locations the from accelerometer data

Table 4. Data Session Charts

Trimmed Data	**Untrimmed Data**
3-axis Magnetometer	3-axis Magnetometer
3-axis Accelerometer	3-axis Accelerometer 10
3-axis Gyroscope	3-axis Gyroscope
Synthesized Magnetometer	Synthesized Magnetometer
Synthesized Accelerometer	Synthesized Accelerometer
Synthesized Gyroscope	Synthesized Gyroscope

stream (5.506 g $\times 10^{-6}$) and gyroscope data stream (1.014 I $\times 10^{-5}$) as well. The start and stop locations generated for the three data streams were correlated to verify their applicability to the sensor session the window algorithm was searching through. Out of the nine entities requiring a window in order to trim off the non-signature portion of the sensor data, seven would have been identifiable with the window algorithm utilizing just the magnetometer data. The exception were the two low decibel subwoofer trials at both 40Hz and 50Hz, these required the accelerometer data for the window algorithm to return an accurate start and stop location.

The window algorithm was built and run in the R language and environment for statistical computing. All preprocessing and statistical processing was performed in R, an effort was made to limit the varieties of software required to replicate this project. Once the start and stop locations were identified for each data session, charts were generated depicting both raw and synthetic sensor readings for the magnetometer, accelerometer, and gyroscope for both the trimmed and full data session, thus each data session resulted in 12 charts (Table 4) being created and stored for visual analysis.

The trimmed data was then used to compute seventy-two statistical attributes for the classifier. The range, standard deviation (SD), variance, skewness, kurtosis, and root-mean-square (RMS) were computed for the magnetometer, accelerometer, and gyroscope, resulting in eighteen attributes. Then, a simple moving average algorithm (SMA) is applied to the data for each sensor's data stream to smooth the data. The SMA algorithm is provided samples sizes of three, five, and seven; the the previous statistical attributes are applied to each of the sensor's 'smoothed' data streams, respectively. This results in fifty-four new attributes, for a total of seventy-two attributes. Table 5 lists the attributes computed for each sensors data.

Table 5. Attributes for the Magnetometer, Accelerometer, and Gyroscope data

Raw Data	**SMA(3)**	**SMA(5)**	**SMA(7)**
Range	Range	Range	Range
SD	SD	SD	SD
Variance	Variance	Variance	Variance
Skewness	Skewness	Skewness	Skewness
Kurtosis	Kurtosis	Kurtosis	Kurtosis
RMS	RMS	RMS	RMS

Additionally, the trimmed data from experiment 2, the vehicles undercarriage scan, was subject to discrete wavelet transformation. The vehicle undercarriage magnetometer signatures had 5 levels of discrete wavelet transforms (DWT) performed. Each level of the DWT results in a set of coefficients that represent the most significant portions of the signal from the previous level of DWT. As such, each level of DWT decomposition will have less coefficients than the level immediately prior. After the coefficients are calculated for each DWT level, the results of levels 2 through 5 will be ordered with a set of the highest and lowest coefficients serving as inputs to the WEKA J4.8 decision tree maker for analysis. Table 6 contains the above details

Table 6. Discrete Wavelet Transform Coefficients

DWT Level	# of High Coefficients	# of Low Coefficients
2	10	10
3	10	10
4	10	10
5	3	3

as well as the number of coefficients being abstracted at each level for inclusion in the J4.8 algorithm. DWT level 5 has less coefficients than the other levels due to the nature of decomposition; there are less coefficients at decomposition level 5 to utilize for analysis.

After the attributes were calculated, the attributes were output into an .arff file for submission to the machine learning workbench WEKA [16] where they were compiled and analyzed using a variety of attributes to determine the best mix for accurate classification of the entities.

IV. Results and Analysis - Recognition and Identification

4.1 Decision Model Review

After gathering the sensor data from the control variables identified in the methodology section, the data was preprocessed, statistically analyzed, and classified via two distinct approaches. The sensor data was exported from the iOS SQLITE database via the SensorSuite program designed specifically for the purpose of acquiring data streams from the available sensors and exporting those sensor streams for aggregation and analysis.

In order to determine the ability of known recognition algorithms to accurately assign an entity to the appropriate classification, two approaches are utilized. Both approaches use the J4.8 WEKA implementation of the C4.5 revision 8 decision tree learner in 14 specific attribute configurations. In the first method, this approach is combined with 10-fold cross-validation and a randomly ordered set of 354 instances to determine the decision tree's ability to accurately classify entities. In the second method, the decision tree learner is paired with separate training and test data sets. The training data set is 254 entities and the test set is 100 entities.

The parameters used to generate the decision tree model via the J4.8 decision tree learner are listed in Table 7.

4.2 Decision Model Results

Generating attributes from the raw data, there are six attributes for each of the four averaging techniques listed in Table 5, thus there are 24 attributes for each of the three sensors yielding a total of 72 attributes for potential inclusion in the decision

Table 7. J4.8 Decision Tree Learner Parameters

Parameter	Value		Parameter	Value
binarySplits	False		saveInstanceData	False
confidenceFactor	0.25		seed	1
debug	False		subtreeRaising	True
minNumObj	2		unpruned	False
numFolds	3		useLaplace	False
reducedErrorPruning	False			

tree modeled by the J4.8 decision tree learner. The attribute sets listed in Table 8 were built and compared for their ability to correctly classify instances.

Table 8. Attribute Set Table

Set #	Mag [a]	Accel [b]	Gyro [c]	Mag [a, d]	Accel [b, d]	Gyro [c, d]
1	X	X	X			
2	X	X				
3	X		X			
4	X					
5		X	X			
6		X				
7			X			
8				X	X	X
9				X	X	
10				X		X
11				X		
12					X	X
13					X	
14						X

[a] Magnetometer abbreviated Mag
[b] Accelerometer abbreviated Accel
[c] Gyroscope abbreviated Gyro
[d] Set does not include SMA attributes

By creating models based off the attribute sets listed in Table 8, results are generated that will help identify which aspects of the multimodal sensor data stream are most useful to recognizing and classifying the entities being investigated. The results demonstrate whether there is value added to fusing data for interrogation. Additionally, the results reveal whether smoothing averages are helpful. Though this last point is harder to prove, as smoothing helps to eliminate outliers in a data stream and there may not be any outliers in the 354 entity instances tested.

The desired result of multimodal sensing would be to accurately classify all entities that a classifier is capable of handling. Careful analysis of the models generated by the J48 from the above attribute sets (Table 8) will reveal that perfect classification is not possible given imperfect data sets. Even those collected in an organized experiment suffer from extraneous data points, unintended effects, and algorithmic imperfection (such as those introduced by trimming the data with the windowing algorithm). More detailed analysis of the correctly classified entities and the incorrectly classified entities will reveal that certain sensors are better at classifying entities that exhibit specific combinations of environmental effects.

For instance the results listed throughout this section will show that for detecting microwaves, the magnetometer and the data stream it outputs are an important tool. When analyzing a subwoofer operating at different frequencies and at different decibel levels, the magnetometer remains important for differentiating between decibel levels, but the accelerometer and gyroscope and their ability to detect rotation and vibration become important for differentiating between frequencies.

There may be no best model for entity detection, therefor it is important to understand what the strengths of a particular model are as well as the effects the entities being classified produce. Some flexibility in model selection would be useful in order to select certain sensor combinations depending on the most likely entity attempting to be classified. As noted in literature review section, [28] proposes a similar system with the JIGSAW algorithm.

Using 10-fold cross validation with each of the attribute sets listed in Table 8, the models obtained the following efficiencies listed in Table 9.

Table 9. 10-Fold Cross-Validation Attribute Set Results

Set #	Correct	Incorrect	Inter-Category	Tree Size	Leaves
1	345	9	4	19	10
2	346	8	4	19	10
3	347	7	3	19	10
4	345	9	2	19	10
5	298	56	4	45	23
6	295	59	4	41	21
7	272	82	8	59	30
8	344	10	3	19	10
9	348	6	0	19	10
10	346	8	4	19	10
11	346	8	0	19	10
12	303	51	4	39	20
13	297	57	4	27	14
14	266	86	6	75	38

[a] Tree size and leaf number vary greatly between certain attribute sets; number of entities remains constant at 10.

From the cross-validation results summary in Table 9, a few commonalities may be ascertained. Overfitting can occur with attributes that offer continuous values for decisions, continuous attributes lend to the construction of decision trees with a high number of branches. With the smallest tree containing 19 nodes, or alternately being of size 19, and the largest tree containing 75 nodes, there is a large disparity in *fit*. Using an average tree size of approximately 30, it is possible to pare to the number of attribute sets into less brittle decision trees. As such, sets 1 through 4 and 8 through 11, and 13 are candidates for further consideration.

When analyzing the results of the candidate attribute sets, Table 9 contains the number of correctly and incorrectly classified entities. Additionally, the number of inter-category classifications if noted. Inter-category refers to misclassifications where the classified entity is placed in the incorrect category, not just sub-category. As such, if a subwoofer is classified as a microwave it is considered inter-category. If a subwoofer operating at 50Hz is classified as a subwoofer operating at 40Hz it is considered intra-category. The goal of entity recognition is to correctly sub-categorize an entity as accurately as possible, however, there is value to being able to categorize an entity into a category even when resolving to the correct sub-category is not possible.

The results of the well-fit decision trees built from attribute sets 1 through 4 all lend to the inclusion of magnetometer data into the classification model. Note that sets 1 through 4 include the SMA data for analysis and should be considered before non-SMA attribute sets if elimination of outlier data via averaging is desired. A quick review of the summarized data leads one to believe that with the control variables utilized in the experiment discussed in section 3, an attribute set based exclusively on magnetometer data is sufficient to categorize the entities correctly. Indeed, the strictly magnetometer set (attribute set 4) posts the most accurate classification results based on inter-category classifications. This is due to the measurable magnetic fields generated by the control variables.

The magnetic field strength as measured by the magnetometer in the smart phone allows for the accurate classification of the entities utilized as control variables. There is enough magnetic difference between microwaves, subwoofers, and the baseline environment to allow for near perfect categorization. Only 2 entities from attribute set 4 are classified outside their respective categories, the other 7 entities are misclassified within their subcategories. The statistical classification methods are accurate enough to determine entity categories for 97.45% of the entity samples. In fact, analysis of the decision tree generated by WEKA (displayed in Figure 2) indicates that the model is able to perform accurate classification using the SD (for the raw data, SMA 3, and SMA 7) and the RMS (for the raw data and SMA 3).

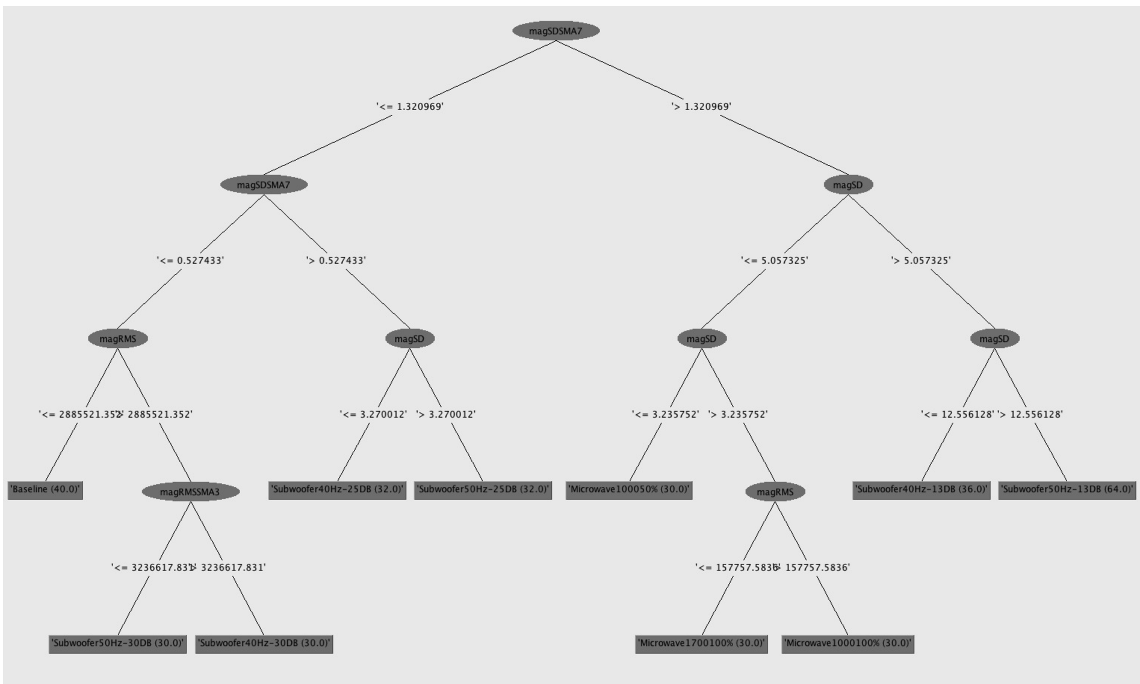


Figure 2. Attribute Set 4 Decision Tree

The confusion matrix (Figure 3) for attribute set 4 shows that two of the subwoofer entities are misclassified as microwave entities, f and i are classified as a and b, respectively. The control variable table for experiment 1 is Table 2 and the entities correlate directly to their confusion matrix letter value. It can be surmised that the magnetic qualities for one of the entity samples for the 40Hz subwoofer and one of the entity samples for the 50Hz subwoofer are similar to the magnetic qualities output by the 1000 watt microwave. This pattern of misclassification is not static through attribute sets 1, 2, 3, and 4. In attribute set 1, there are four category level misclassifications, two in the microwave category (one each on the 1000 watt microwave (b) and the 1600 watt microwave (c)), one in the h level control variable entity (50Hz subwoofer) and another in the (j) level control variable entity (baseline). In attribute set 2, there are four category level misclassifications, one at the microwave level (c), one at the subwoofer category level (h), and two at the (j) level control variable entity (baseline). In attribute set 3, there are three category level misclassifications, two in the subwoofer category level (one each at 40Hz (f) and 50Hz (i)) and another in the (j) level control variable entity (baseline). A review of attribute sets 1 through 3 helps reveal what decision tree nodes are offered compared to the previously discussed attribute set 4, and can offer an intuition as to why the misclassified entities changes between models.

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g  h  i  j   <-- classified as
 30  0  0  0  0  0  0  0  0  0 |  a = Microwave100050%
  0 30  0  0  0  0  0  0  0  0 |  b = Microwave1000100%
  0  2 28  0  0  0  0  0  0  0 |  c = Microwave1700100%
  0  0  0 35  0  0  1  0  0  0 |  d = Subwoofer40Hz-13DB
  0  0  0  0 29  1  0  0  0  0 |  e = Subwoofer40Hz-30DB
  1  0  0  0  0 29  0  0  2  0 |  f = Subwoofer40Hz-25DB
  0  0  0  0  0  0 64  0  0  0 |  g = Subwoofer50Hz-13DB
  0  0  0  0  1  0  0 29  0  0 |  h = Subwoofer50Hz-30DB
  0  1  0  0  0  0  0  0 31  0 |  i = Subwoofer50Hz-25DB
  0  0  0  0  0  0  0  0  0 40 |  j = Baseline
```

Figure 3. Attribute Set 4 Confusion Matrix

Within the SMA attributes sets, the set with the magnetometer and gyroscope (attribute set 3) offer the lowest number of incorrectly classified entities. The decision tree includes 5 magnetometer attribute nodes and 4 gyroscope attribute nodes. With 7 misclassifications, set 3 offers 2 more correct classifications than set 4 for an accuracy rate of 98.02%. However, there is 1 additional inter-category misclassification. While overall sub-level classification has improved, classification in the parent categories has worsened. Instead of relying strictly on statistical values based on magnetic properties, the model built from attribute set 3 includes statistical values based on the gyroscope in addition to the magnetometer values. This allows the model to more accurately classify the subwoofer entities via the subwoofer sound wave and the torque experienced by the smart phone from the sound wave. The addition of torque nodes in the decision tree introduces a misclassified baseline reading that was classified correctly when just magnetic statistical methods were utilized as in set 4. Additionally, the the two subwoofer entities that were misclassified in attribute set 4 remain in the set 3 confusion matrix. This demonstrates how a multimodal approach offers both addition resolution possibilities as well as potential misclassifications due to similarity in statistical decision node values.

Attribute set 2 decreases in accuracy by one additional misclassification, as well as one additional inter-category misclassification, over attribute set 3. The decision tree includes 5 accelerometer attribute nodes and 4 magnetometer attribute nodes. The accuracy rate of set 2 is 97.74%. This is still better than the 9 misclassifications offered by attribute set 4, but worse than the inter-category misclassification rate of 2 for set 4. The inclusion of accelerometer data helped eliminate the misclassified subwoofer entries present in sets 3 and 4. However, the loss of magnetometer based decision nodes increase the baseline misclassifications to 2, as well as introduce a microwave misclassified as a subwoofer. Lastly, a new misclassification appears in the subwoofers where 1 control variable (h) is classified as a baseline reading. Attribute set 2 continues to demonstrate how a particular model could be tuned to certain types of entities, in this case those that induce movement on a sensor platform.

Attribute set 1 contains the magnetometer, accelerometer, and gyroscope attributes, as well as their SMA statistical attributes. The number of misclassifications in 9 is similar to attribute set 4, but with 4 inter-category misclassifications, this makes attribute set 1 97.45% accurate. The model produced by the J48 decision tree maker contains 2 accelerometer attribute nodes, 3 gyroscope attribute nodes, and 3 magnetometer attribute nodes. The inclusion of all 3 sensor attribute sets results in the misclassification of 2 microwave control variables, both as subwoofer. Additionally, the subwoofer from set 2 classified as a baseline entity is now classified as a microwave. Lastly, there is a baseline reading classified as a subwoofer, indicating that the lack of magnetometer decision points is impacting baseline classification. Figure 4 contains the confusion matrices for attribute sets 1 through 4.

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g  h  i  j   <-- classified as
 30  0  0  0  0  0  0  0  0  0 |  a = Microwave100050%
  0 29  0  0  0  0  0  1  0  0 |  b = Microwave1000100%
  0  1 28  0  1  0  0  0  0  0 |  c = Microwave1700100%
  0  0  0 36  0  0  0  0  0  0 |  d = Subwoofer40Hz-13DB
  0  0  0  0 30  0  0  0  0  0 |  e = Subwoofer40Hz-30DB
  0  0  0  0  0 31  0  0  1  0 |  f = Subwoofer40Hz-25DB
  0  0  0  0  0  1 62  0  1  0 |  g = Subwoofer50Hz-13DB
  1  0  0  0  0  0  0 28  1  0 |  h = Subwoofer50Hz-30DB
  0  0  0  0  0  0  0  0 32  0 |  i = Subwoofer50Hz-25DB
  0  0  0  0  0  0  0  1  0 39 |  j = Baseline
```

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g  h  i  j   <-- classified as
 30  0  0  0  0  0  0  0  0  0 |  a = Microwave100050%
  0 30  0  0  0  0  0  0  0  0 |  b = Microwave1000100%
  0  1 28  0  0  0  0  1  0  0 |  c = Microwave1700100%
  0  0  0 35  0  0  1  0  0  0 |  d = Subwoofer40Hz-13DB
  0  0  0  0 30  0  0  0  0  0 |  e = Subwoofer40Hz-30DB
  0  0  0  0  0 32  0  0  0  0 |  f = Subwoofer40Hz-25DB
  0  0  0  0  0  0 64  0  0  0 |  g = Subwoofer50Hz-13DB
  0  0  0  0  1  0  0 27  1  1 |  h = Subwoofer50Hz-30DB
  0  0  0  0  0  0  0  0 32  0 |  i = Subwoofer50Hz-25DB
  1  0  0  0  0  0  0  1  0 38 |  j = Baseline
```

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g  h  i  j   <-- classified as
 30  0  0  0  0  0  0  0  0  0 |  a = Microwave100050%
  0 30  0  0  0  0  0  0  0  0 |  b = Microwave1000100%
  0  2 28  0  0  0  0  0  0  0 |  c = Microwave1700100%
  0  0  0 36  0  0  0  0  0  0 |  d = Subwoofer40Hz-13DB
  0  0  0  0 29  0  0  0  1  0 |  e = Subwoofer40Hz-30DB
  1  0  0  0  0 30  0  0  1  0 |  f = Subwoofer40Hz-25DB
  0  0  0  0  0  0 64  0  0  0 |  g = Subwoofer50Hz-13DB
  0  0  0  0  0  0  0 30  0  0 |  h = Subwoofer50Hz-30DB
  0  1  0  0  0  0  0  0 31  0 |  i = Subwoofer50Hz-25DB
  0  0  0  0  0  0  0  1  0 39 |  j = Baseline
```

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g  h  i  j   <-- classified as
 30  0  0  0  0  0  0  0  0  0 |  a = Microwave100050%
  0 30  0  0  0  0  0  0  0  0 |  b = Microwave1000100%
  0  2 28  0  0  0  0  0  0  0 |  c = Microwave1700100%
  0  0  0 35  0  0  1  0  0  0 |  d = Subwoofer40Hz-13DB
  0  0  0  0 29  1  0  0  0  0 |  e = Subwoofer40Hz-30DB
  1  0  0  0  0 29  0  0  2  0 |  f = Subwoofer40Hz-25DB
  0  0  0  0  0  0 64  0  0  0 |  g = Subwoofer50Hz-13DB
  0  0  0  0  1  0  0 29  0  0 |  h = Subwoofer50Hz-30DB
  0  1  0  0  0  0  0  0 31  0 |  i = Subwoofer50Hz-25DB
  0  0  0  0  0  0  0  0  0 40 |  j = Baseline
```

Attribute Set 1 (Top-Left), Attribute Set 2 (Top-Right), Attribute Set 3 (Bottom-Left), and Attribute Set 4 (Bottom-Right)

Figure 4. Attribute Sets 1 - 4 Confusion Matrices

A review of the attribute sets that do not include SMA values includes attribute sets 8 through 11 and attribute set 13. The model generated for attribute set 13 is 42% larger than the models for attribute sets 1 through 4, and as such may be overfit. Attributes sets 1 through 7 are the SMA versions and directly correlate to the non-SMA attribute sets 8 through 14, respectively. As such, the expectation that both attribute sets 4 and 11, the strictly magnetometer attribute sets, would perform similarly is upheld. Attribute set 11 performs slightly better than attribute set 4, with 1 additional correct classification and zero inter-category classifications. The results of attribute set 9 compared to set 2 are likewise similar. Attribute set 9 has 2 additional correct classifications and has zero inter-category classifications, demonstrating that the pairing of the accelerometer and magnetometer can produce highly accurate categorical classification results.

The decision tree for attribute set 13 is tree of size 27, which as noted previously is 42% larger than the decision tree's of size 19 for the previously discussed attribute sets. While this model is closer to overfit than the previous models, it is not as egregiously overfit as the other non-magnetometer based attribute sets. Attribute set 13 is based strictly on non-SMA accelerometer data and reveals the possibility of constructing decision trees based off sensors other than a magnetometer for entity detection, recognition, and classification. Once again, this sheds light on the need to construct a decision tree making algorithm that is geared towards a particular set of categories eligible for detection.

The fact that the non-SMA attribute sets produced results very similar to the SMA attribute sets signifies a reliability to the training and testing data that minimizes the need to average data samples to eliminate noise. A real world application of entity sensing would include spikes in various sensor data that may not be indicative of an entities existence. As such the SMA attribute sets, as seen in the literature for activity recognition, would probably be more applicable in a non-experiment driven entity classification scheme. The non-SMA attribute sets would include peaks and troughs that may make accurate classification more difficult.

4.3 Training and Test Set Review

As a compliment to the 10-fold cross validation reviewed in Section 4.1, a model was built from a training set of 254 instances. A test set of 100 instances was run with each of the attribute sets listed in Table 8, the models obtained the following efficiencies listed in Table 10. Comparison between Table 10 and Table 9 shows that with little exception, the models created for the attribute sets listed in Table 8 are similar between the two model creation methods. This is to be expected as the J4.8 is utilized to generate both sets of models, the only major difference is the size of the fold, as the holding back of a 100 entity test set is approximately 28% of the set, versus 10% in the 10-fold cross validation methodology.

Table 10. Training and Test Model Attribute Set Results

Set #	Correct[b]	Incorrect[b]	Inter-Category[b]	Tree Size	Leaves
1	99	1	1	19	10
2	99	1	1	19	10
3	99	1	1	19	10
4	99	1	1	19	10
5	84	16	1	39	20
6	78	22	1	37	19
7	73	27	4	41	21
8	98	2	2	19	10
9	98	2	2	19	10
10	100	0	0	19	10
11	95	5	0	19	10
12	84	16	2	43	22
13	85	15	2	47	24
14	74	26	2	51	26

[a] Tree size and leaf number vary greatly between certain attribute sets; number of entities remains constant at 10.
[b] Classification accuracy based on test set results. Training set consists of 254 entity instances, test set consists of 100 entitty instances.

4.4 Graph Analysis

Visual analysis of a data session from a random instance of each of the control variables provides insight into which sensors are useful for detecting a particular entity. While a visual analysis may provide the insight into sensor selection, it is not a substitute for statistical analysis due to subtle changes the sensor may detect. The plots in Figures 5, 6, and 7 are on a time scale (x-axis) where the sensor data from each of the respective sensors 3 axises have been normalized by synthesization (y-axis) where

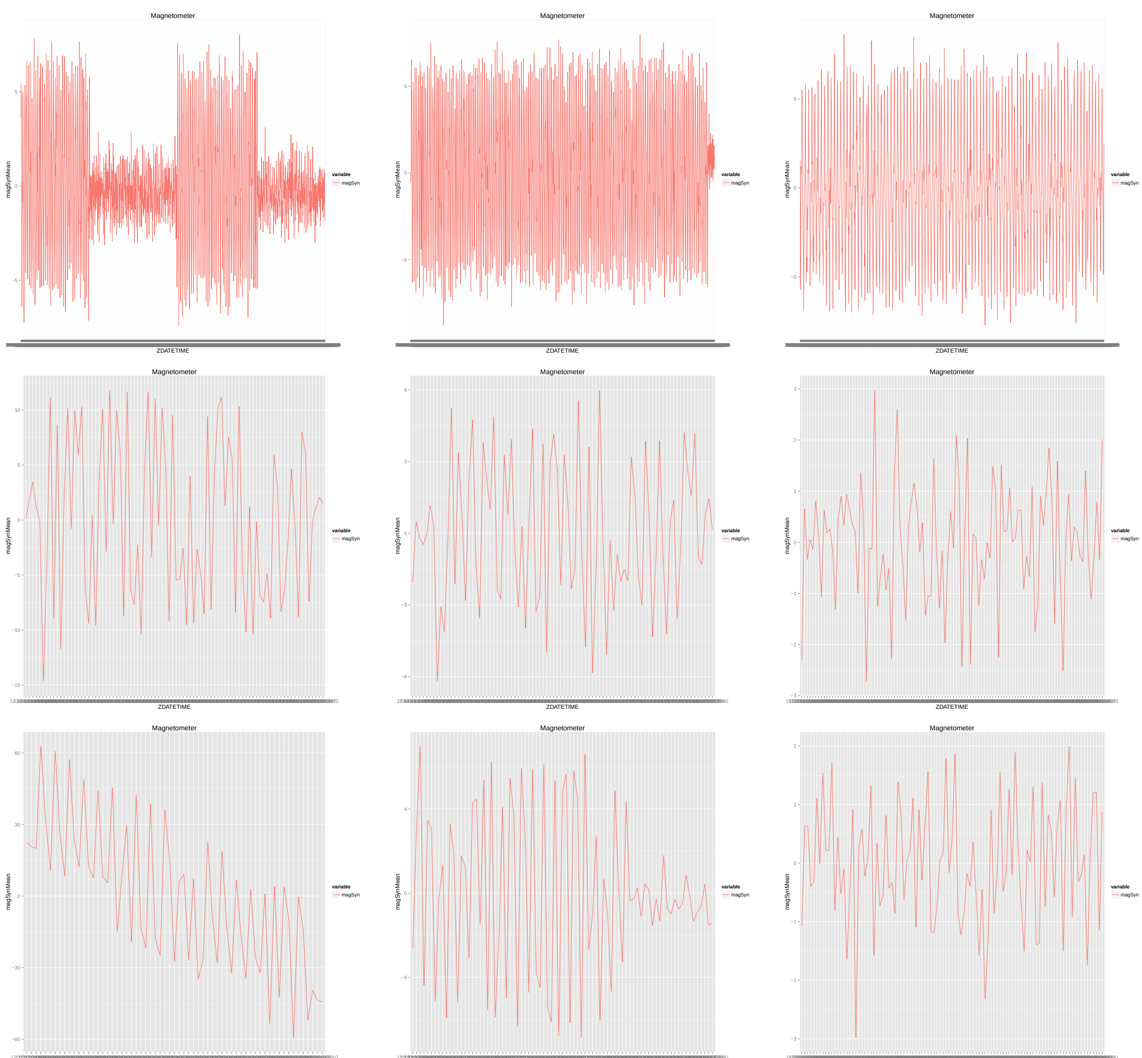
$$A_{sensor} = \sqrt{(A_x)^2 + (A_y)^2 + (A_z)^2}$$

Thus graphs depict the orientation independent effects experienced by the smart phone sensors.

The magnetometer charts in Figure 5 provide evidence that each of the control variables produce different magnetic effects. Some of these effects are quite apparent visually, such as the difference between the 1000 watt microwave operating at 50% power and 100% power. Others are less apparent, though still present, such as the change in range exhibited between various subwoofer dB level. For the subwoofer, the louder dB levels (the larger dB values) produce a more measurable magnetic field. This will be revealed in the analysis of the statistical data in Tables 11, 12, and 13.

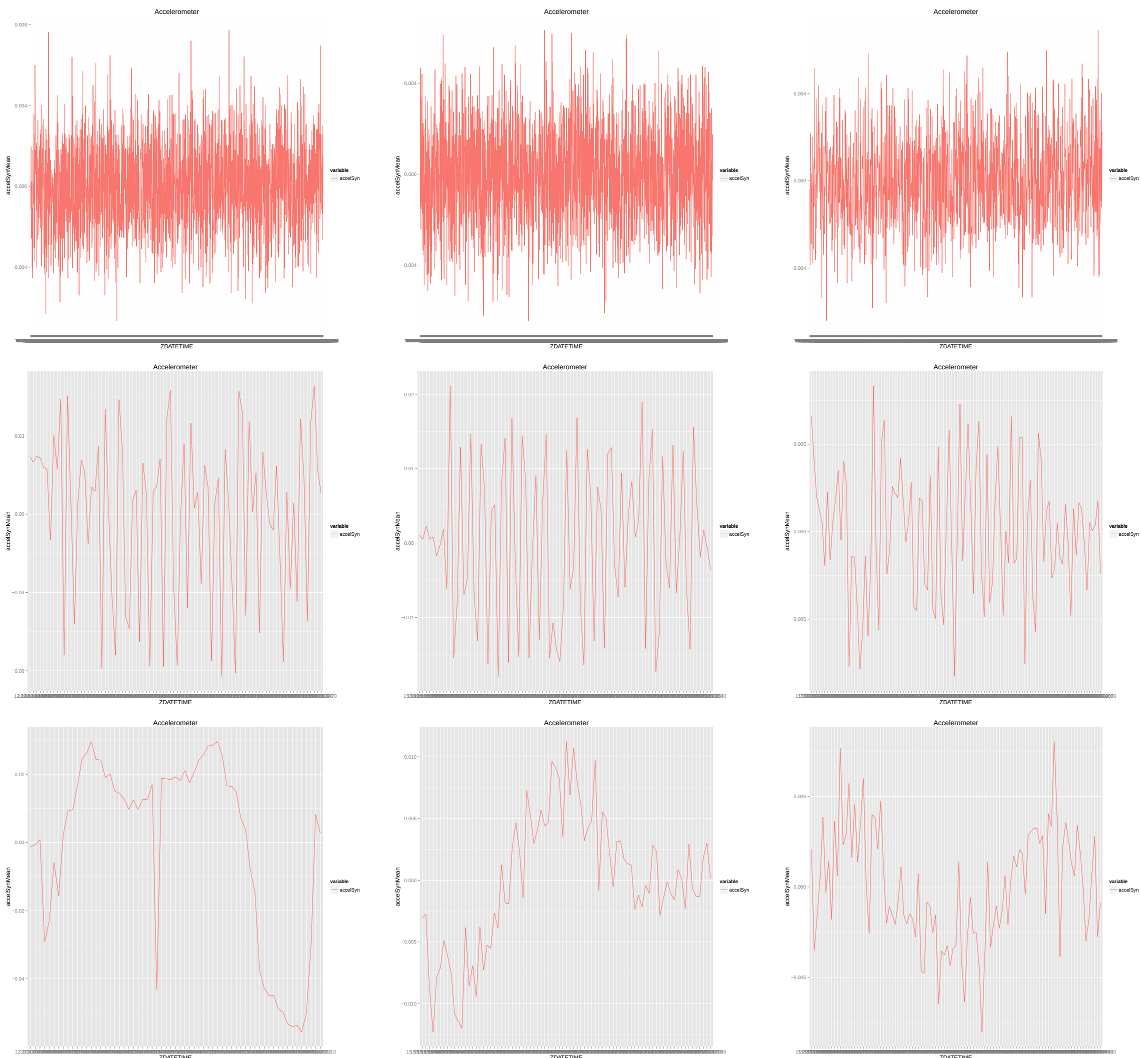
The accelerometer charts in Figure 6 provide evidence that some of the control variables produce seemingly different gravitational effects. Each of the microwaves have a fairly similar accelerometer data range. However, each of the subwoofer control variables appear different. With decreasing dB level, the gravitational effects due to sound wave vibration decrease. This remains consistent between the two frequency levels. Additionally, sampling rate seems to play a role when comparing the 40Hz and 50Hz output, presumably due to the nyquist interval.

The gyroscope charts in Figure 7 reveal similar environmental aspects to the accelerometer charts in Figure 6. An entity producing torque measurable effects is



Microwave (Top Row) at 1000 Watt at 50% Power (Left), at 1000 Watt at 100% Power(Center), at 1600 Watt at 100% Power(Right), 40Hz Subwoofer (Middle Row) at -13dB (Left), at -25dB (Center), at -30dB(Right), 50Hz Subwoofer (Bottom Row) at -13dB(Left), at -25dB(Center), and at -30dB(Right)

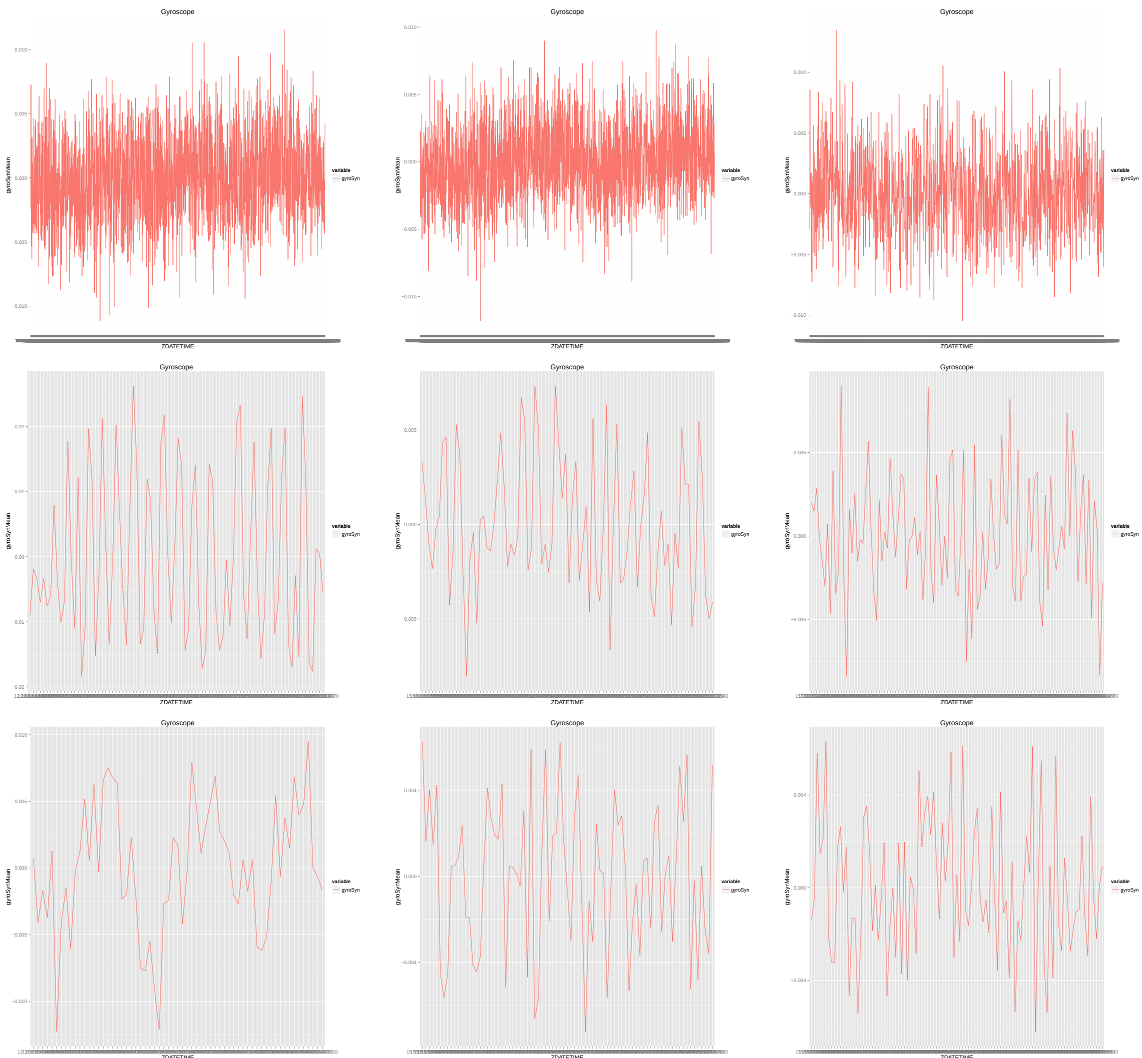
Figure 5. Randomly Selected Control Variable Magnetometer Data



Microwave (Top Row) at 1000 Watt at 50% Power (Left), at 1000 Watt at 100% Power(Center), at 1600 Watt at 100% Power(Right), 40Hz Subwoofer (Middle Row) at -13dB (Left), at -25dB (Center), at -30dB(Right), 50Hz Subwoofer (Bottom Row) at -13dB(Left), at -25dB(Center), and at -30dB(Right)

Figure 6. Randomly Selected Control Variable Accelerometer Data

likely to also be producing vibrational effects that are measurable by the accelerometer. Thus it is no surprise that the accelerometer and gyroscope charts are visually similar.



Microwave (Top Row) at 1000 Watt at 50% Power (Left), at 1000 Watt at 100% Power(Center), at 1600 Watt at 100% Power(Right), 40Hz Subwoofer (Middle Row) at -13dB (Left), at -25dB (Center), at -30dB(Right), 50Hz Subwoofer (Bottom Row) at -13dB(Left), at -25dB(Center), and at -30dB(Right)

Figure 7. Randomly Selected Control Variable Gyroscope Data

4.5 Statistical Analysis

Analyzing the mean of the statistical values produced for each of the control variables categories helps to understand what qualities in the graphs in Figures 5, 6, and 7 are useful for classification purposes. Included are the range, SD, variance, skewness, kurtosis, and the RMS. Range represents the difference between high and low readings in the sensor data. SD represents the standard deviation and is the amount of variation from average. Variance is the measure of how far a set of numbers is spread out. Skewness is a measure of the asymmetry of the probability distribution. Kurtosis is the measure of the peakedness of a probability distribution, as such it is a probability distribution shape descriptor like skewness. RMS measures the magnitude of the data stream.

Table 11. Magnetometer Statistical Values for Control Variables

Control Variable	Range[a]	SD[a]	Variance[a]	Skewness[a]	Kurtosis[a]	RMS[a]
a	19.857	2.866	8.225	0.109	3.878	176750
b	21.425	3.82	14.614	0.085	2.547	176860
c	18.009	3.861	14.917	0.184	2.493	157474
d	39.536	11.93	143.018	0.061	1.665	2705813
f	11.748	3.045	9.289	0.009	1.979	3303501
e	5.515	1.121	1.264	-0.095	2.789	3302969
g	62.17	19.194	386.335	0.121	1.67	2090392
i	15.567	4.445	19.919	0.053	1.785	3302787
h	5.186	1.051	1.111	0.075	2.908	3297486
j	5.484	0.940	0.885	-0.013	2.909	1005972

[a] All values were read from WEKA Explorer's attribute section and are limited to 3 significant digits.

The magnetometer's statistical output (Table 11) helps illuminate some of the pertinent details for the control variables. For instance, when attempting to categorize the 1000 watt microwave in 100% power and 50% power, the range may not offer enough statistical difference to be useful. The variance on the other hand offers a more appealing attribute for classification purposes. Given that the J4.8 decision tree maker creates nodes for classification purposes, the classifier may not always choose the **best** attribute for classification, but it will choose an attribute that helps classify an entity. What this means is that even though the difference in magnetic variance seems to be a clear choice for differentiating between a 1000 watt microwave operating at 2 different power settings, the classifier utilizes the Standard Deviation. This doesn't make the classifier wrong in any sense, it just shows that one's intuition may not be the same as the decision maker's algorithms. T-Tests performed on the attributes chosen for the decision tree verify the statistical validity of the J4.8 decision tree maker's choices (Table 14).

Within the magnetometer data there is a relatively narrow range that the statistical values for microwaves exhibit. The differences are enough for accurate classification and the decision tree does well classifying the microwave correctly. For the subwoofer, both at the 40Hz and 50Hz frequency, there are significant differences in the magnetometer values between dB levels. This confirms the effects seen on the graphs in Figure 5, as the magnetic properties being measured are directly correlated to the dB level. The larger the dB, the larger the magnetic field generated by the subwoofer. This effect is detected in range, SD, and variance most pronouncedly. The lowest dB level subwoofer attributes are similar to those seen in the baseline statistics, indicating that magnetic values may not reveal the presence of a quiet subwoofer.

Analyzing the accelerometer statistical values in Table 12 helps to identify when accelerometer data may be useful for categorization. Since the decision tree generated by the J4.8 was of size 27 for our attribute set of strictly raw accelerometer data, it is a fairly brittle decision tree that would benefit from the inclusion of another sensor's attributes.

The raw values from the accelerometer are a function of gravity, with a purely stationary data reading being a value of 1.0 for the force of gravity. As the strength of gravity varies in many different ways, proximity to other objects, elevation, and latitudinal position on the earth, a reading of 1.0 should not be expected. As such the apparent variance in readings experienced by a *near* stationary smart phone that will be on the order 3 to 6 significant digits smaller than 1.0. Combined with the

Table 12. Accelerometer Statistical Values for Control Variables

Control Variable	Range[a]	SD[a]	Variance[a]	Skewness[a]	Kurtosis[a]	RMS[a]
a	0.014	0.002	0[b]	-0.003	2.986	0.952
b	0.014	0.002	0[b]	-0.009	3.023	0.954
c	0.017	0.002	0[b]	-0.128	4.419	0.955
d	0.069	0.021	0[b]	-0.137	1.607	0.939
f	0.038	0.011	0[b]	-0.039	1.728	0.950
e	0.017	0.004	0[b]	0.063	2.402	0.948
g	0.059	0.018	0[b]	-0.170	1.788	0.935
i	0.022	0.005	0[b]	-0.088	2.110	0.947
h	0.013	0.003	0[b]	0.061	2.740	0.941
j	0.012	0.002	0[b]	0.030	2.993	0.938

[a] All values were read from WEKA Explorer's attribute section and are limited to 3 significant digits.
[b] Value not 0, see note a.

manner in which WEKA displays the attribute values in the Explorer tab, the small differences in SD and variance in Table 12 are not evident. What is evident is that the subwoofer produces a sound wave that vibrates the smart phone in a measurable manner. The dB level isn't the only aspect effecting the gravitational field read by the accelerometer, the frequency also effects the sensor readings. The accelerometer range reveals that a larger dB causes the smart phone to register changes in gravitational force. Indeed, with a baseline reading of 0.012 (control variable j), even the microwave ovens effect the environment to some degree (control variables a,b, and c).

The values output by the gyroscope are different from either the magnetometer and accelerometer output. The magnetometer and accelerometer are measures of physical properties that are always present on earth and are easy for us to comprehend. The gyroscope measures the amount of rotation being experienced by the phone and as such is effected to some degree by the rotation of the earth.

The gyroscope's statistical values are listed in Table 13 and continue to illustrate a few of the points highlighted in Tables 11 and 12. First, the gyroscope measures

Table 13. Gyroscope Statistical Values for Control Variables

Control Variable	Range[a]	SD[a]	Variance[a]	Skewness[a]	Kurtosis[a]	RMS[a]
a	0.019	0.003	0[b]	0.095	3.038	0.004
b	0.019	0.003	0[b]	0.083	3.098	0.004
c	0.017	0.003	0[b]	0.064	3.074	0.004
d	0.051	0.014	0[b]	0.336	1.964	0.004
f	0.016	0.003	0[b]	0.037	2.822	0.003
e	0.017	0.003	0[b]	0.107	2.895	0.003
g	0.022	0.005	0[b]	0.140	2.620	0.003
i	0.017	0.003	0[b]	0.106	2.844	0.003
h	0.016	0.003	0[b]	0.215	3.009	0.003
j	0.017	0.003	0[b]	0.110	3.052	0.002

[a] All values were read from WEKA Explorer's attribute section and are limited to 3 significant digits.
[b] Value not 0, see note a.

readily visible differences between the loudest dB level for each frequency and the two quieter dB levels. This shows that with a loud enough sound wave, not only does the smart phone experience gravitational changes related to vibrations, the phone itself is rotating to some degree.

The subwoofer entities with the lowest dB level do not produce a magnetic field that is readily discernible in the bottom-right of Figure 5, however, the effects are pronounced enough that both the accelerometer and gyroscope display a feature in the bottom-right of their respective graphs in Figures 6 and 7. Combined with the decision tree built by WEKA that utilizes the gyroscopes RMS value to determine between baseline and non-baseline entities, there is a strong case to including gyroscope output in an entity recognition algorithm.

4.6 Statistical Analysis of Attribute Set 9

In order to prove that the decision nodes chosen for the decision tree are statistically valid, the nodes in the most accurate attribute set were chosen for analysis. The decision tree generated from attribute set 9 resulted in 6 misclassifications, with zero inter-category misclassifications. Figure 8 shows the decision tree generated by WEKA.

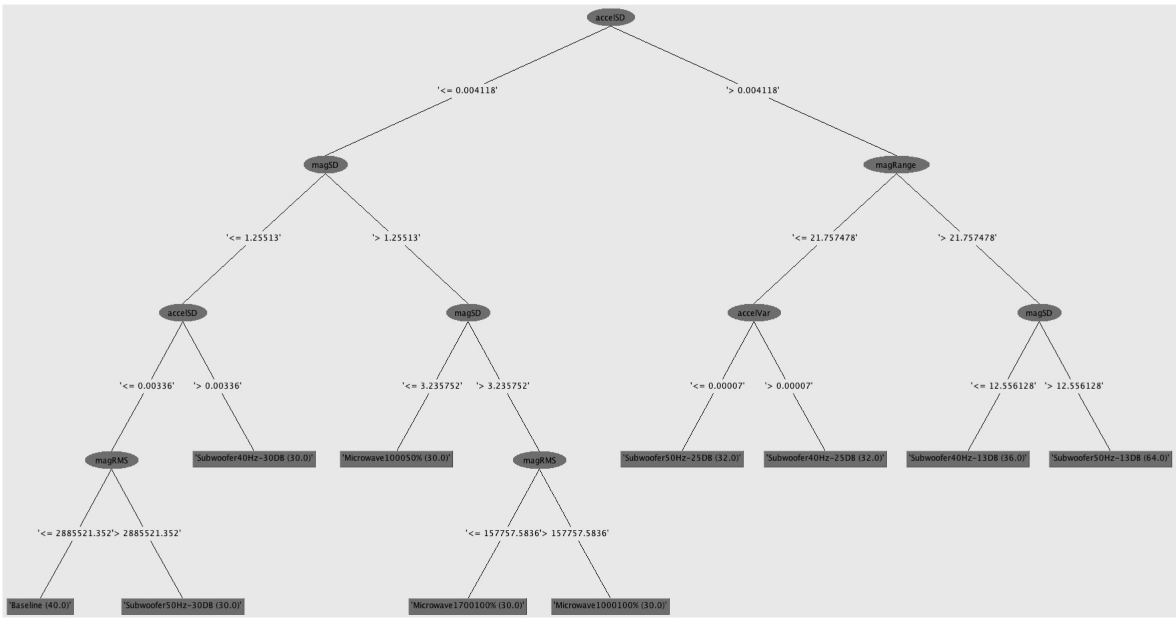


Figure 8. Attribute Sets 9 Decision Tree

Utilizing the decision nodes shown in Figure 8, t-tests were performed on all the nodes. Table 14 shows the results of the t-tests performed in R on the raw statistical attributes generated from the entity data sessions. With very small p-values, the results indicate the validity of using these attributes as decision nodes in the decision tree. As this analysis was performed in R, it is capable of representing far more significant digits than the values shown in Tables 12 and 13.

Table 14. Attribute Set 9 T-Test Results

Node Entities	Attribute	t-value	d.f.[a]	p-value
j & h	Mag RMS	13.5793	39.00	2.30×10^{-16}
c & b	Mag RMS	195.9860	51.16	2.98×10^{-75}
(h, j) & e	Accel SD	26.6304	93.73	1.68×10^{-45}
(a, b, c, e, h, j) & (d, f, g, i)	Accel SD	19.1605	164.87	8.56×10^{-44}
(b, c) & a	Mag SD	38.8817	72.03	4.40×10^{-50}
(a, b, c) & (e, h, j)	Mag SD	48.4696	95.69	3.80×10^{-69}
d & g	Mag SD	13.1731	71.50	8.38×10^{-21}
f & i	Accel Var	53.0371	38.30	1.83×10^{-37}
d & g	Mag Range	23.1603	104.01	7.54×10^{-43}

[a] d.f. stands for degrees of freedom

V. Results and Analysis - Scanning

5.1 Experiment 2

After gathering the sensor data from the control variables for experiment 2 (Table 3) identified in the methodology section, the data was preprocessed, statistically analyzed, transformed, and classified via two distinct approaches. The sensor data was exported from the iOS SQLITE database via the SensorSuite program designed specifically for the purpose of acquiring data streams from the available sensors and exporting those sensor streams for aggregation and analysis.

5.2 Statistical Attributes

Each vehicle's undercarriage was scanned 30 times by the methodology explained in Section 3.2. In addition, 30 baseline readings with no vehicle present were taken as well. Thus there are 90 entities between the 2 vehicles and baseline for control variables. The attribute sets utilized in the statistical analysis are the same as those listed in Table 5.

The results of 10-fold cross-validation on the vehicle's undercarriage experiment show that it is possible to correctly identify between the two vehicles utilized as control variables. The best results with accuracy rates of 96.67% are obtained with attribute sets 1 through 4 and 8 through 11, which are the attribute sets that contain the magnetometer data. The decision tree on the left in Figure 9 is generated from all 72 possible attributes listed in Table 5. Of the 72 possible attributes, only 2 attributes are utilized by the decision tree. In attribute sets that contain both magnetometer and gyroscope attributes, the decision tree maker generated trees that contain a gyroscope RMS attribute (either raw or from SMA3) and the magnetometer's raw range attribute. In the decision tree on the right in Figure 9 the only attribute used is the magnetometer's raw range attribute. An additional decision tree was made with purely SMA based magnetometer data (attribute set 15) to determine whether classification would improve with smoothed data. This yielded the best results overall with 2 incorrectly classified instances, using only the magnetometer's SMA3 range attribute; the decision tree resulted in 1 inter-category misclassification.

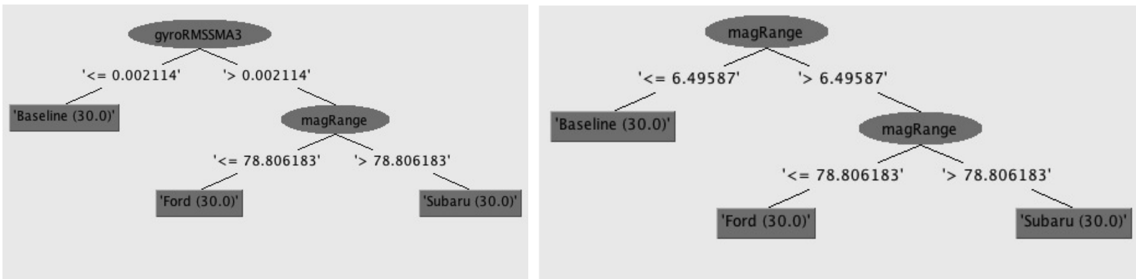


Figure 9. Experiment 2 Decision Trees

From the decision trees generated for Table 15, the typical confusion matrix (attribute sets 1 through 4 and 8 through 11) is shown in Figure 10. The confusion matrix for attribute set 15 contains 1 additional correct classification for the Ford.

Attribute sets that do not include magnetometer data either became very brittle decision trees where the tree was overfit or the results were not as accurate as those including magnetometer attributes. Attribute set 5 managed to produce a satisfactory number of correct classifications, at an accuracy rate of 91.11% with SMA based accelerometer and gyroscope attributes. With a total of 6 leaves and three control variables to classify between, attribute set 5 is probably overfit.

```
=== Confusion Matrix ===

  a  b  c   <-- classified as
 28  2  0 |  a = Ford
  0 30  0 |  b = Subaru
  1  0 29 |  c = Baseline
```

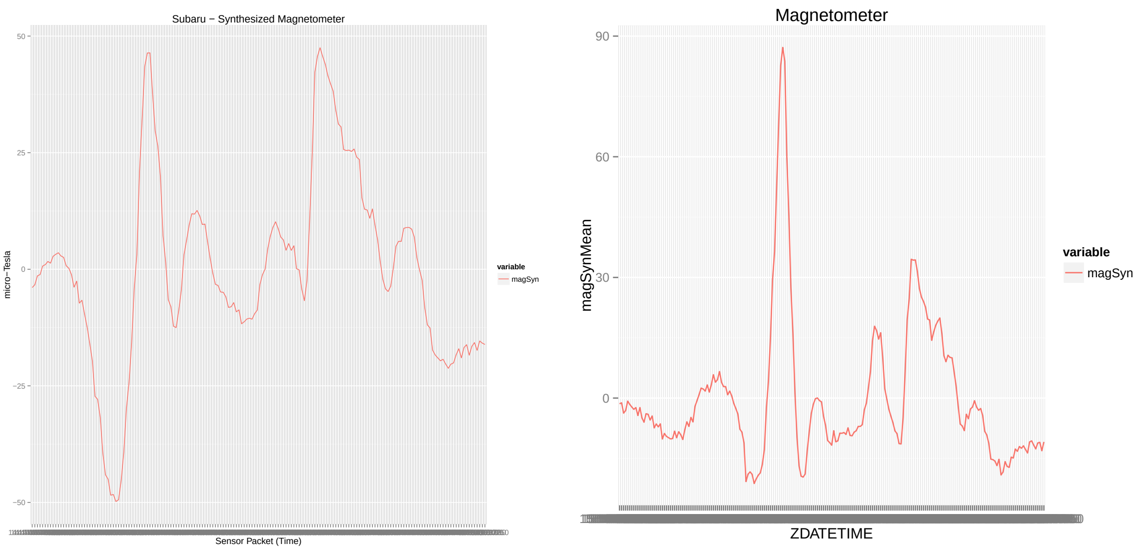
Figure 10. Attribute Set Confusion Matrix For Experiment 2

Table 15. Vehicle 10-Fold Cross-Validation Attribute Set Results

Set #	Correct	Incorrect	Inter-Category	Tree Size	Leaves
1	87	3	1	5	3
2	87	3	1	5	3
3	87	3	1	5	3
4	87	3	1	5	3
5	82	8	1	7	4
6	82	8	2	11	6
7	68	22	1	11	6
8	87	3	1	5	3
9	87	3	1	5	3
10	87	3	1	5	3
11	87	3	1	5	3
12	84	6	1	11	6
13	64	26	15	19	10
14	80	10	1	9	5
15[b]	88	2	1	5	3

[a] Tree size and leaf number vary greatly between certain attribute sets; number of entities remains constant at 10.
[b] SMA based magnetometer attributes.

The results from the attribute sets utilized in experiment 2 demonstrate that the magnetometer is the most valuable sensor for this type of entity classification. The vehicle's undercarriage effects the magnetic field sensed by the magnetometer in a manner significant enough that distinguishing between two different vehicles is possibles with just magnetic attributes. The typical vehicle signatures for both control variables is shown in Figure **??**. When presented the ability to build a decision tree from all available attributes, the J4.8 utilizes a gyroscope attribute that distinguishes between the presence of a vehicle control variable and the baseline readings. This gyroscope attribute alludes to the presence of detectable motion being experienced by the smart phone, however, the lack of gyroscope attributes results in a purely magnetometer based decision tree that is just as accurate for this experiment.



2013 Subaru Crosstrek Undercarriage (Left) and 2013 Ford F-150 (Right)

Figure 11. Example Vehicle Signatures

5.3 Wavelet Decomposition

Another technique for classifying time domain signals is to utilize wavelets. As the signatures shown in Figure **??** demonstrate the presence of peaks and troughs in the magnetometer data, wavelet decomposition offers the possibility of capturing coefficients that are relevant to distinguishing between multiple signatures. Discussed in Section 3.8, wavelet decomposition was performed at the levels noted in Table 6 to obtain the referenced number of high and low coefficients. The results of the wavelet decomposition are noted in Table 16.

Table 16. Vehicle 10-Fold Cross-Validation Coefficient Results

Level[a]	Correct	Incorrect	Inter-Category	Tree Size	Leaves
2	76	14	2	9	5
3	81	9	1	7	4
4	89	1	1	7	4
5	73	17	2	7	4

[a] Decomposition level

By utilizing DWT, the decision tree maker was given the ability to build a decision tree off of signals based more about the control variable's magnetic characteristic representation rather than based purely off the control variable's statistical attributes. Thus the magnetometer's captured signal with the distinctive peaks and troughs were not lost during analysis and multiple occurrences of each were presented to the J4.8. The results demonstrate the ability of the DWT to present coefficients to the decision tree maker that result in highly accurate inter-category classification and depending on the decomposition level, highly accurate overall classification rates.

The best overall results between both the statistical attribute decision trees and the DWT coefficient decision tree are found at the fourth level of decomposition. The results indicate an ability to correctly classify the control variables 98.89% of the time. Using magnetometer data from a data capture of the Subaru undercarriage, decomposition levels 1 through 4 can be seen in Figure 12. The decomposition was performed in R with the wavelets package utilizing the DWT function with levels set to 5, boundary set to reflection, and fast set to false.

The ability to classify the vehicle signatures produced by the control variables in experiment 2 provides for the possibility of expanding the set of classified vehicles to a much larger set. By utilizing signal decomposition coefficients, it opens up the ability to analyze the specific location of peaks and troughs in a signature, allowing an algorithm to discriminate between different vehicles. As construction styles vary, and component placement differs between makes and models, the ability to classify a larger set of vehicles requires additional attention.

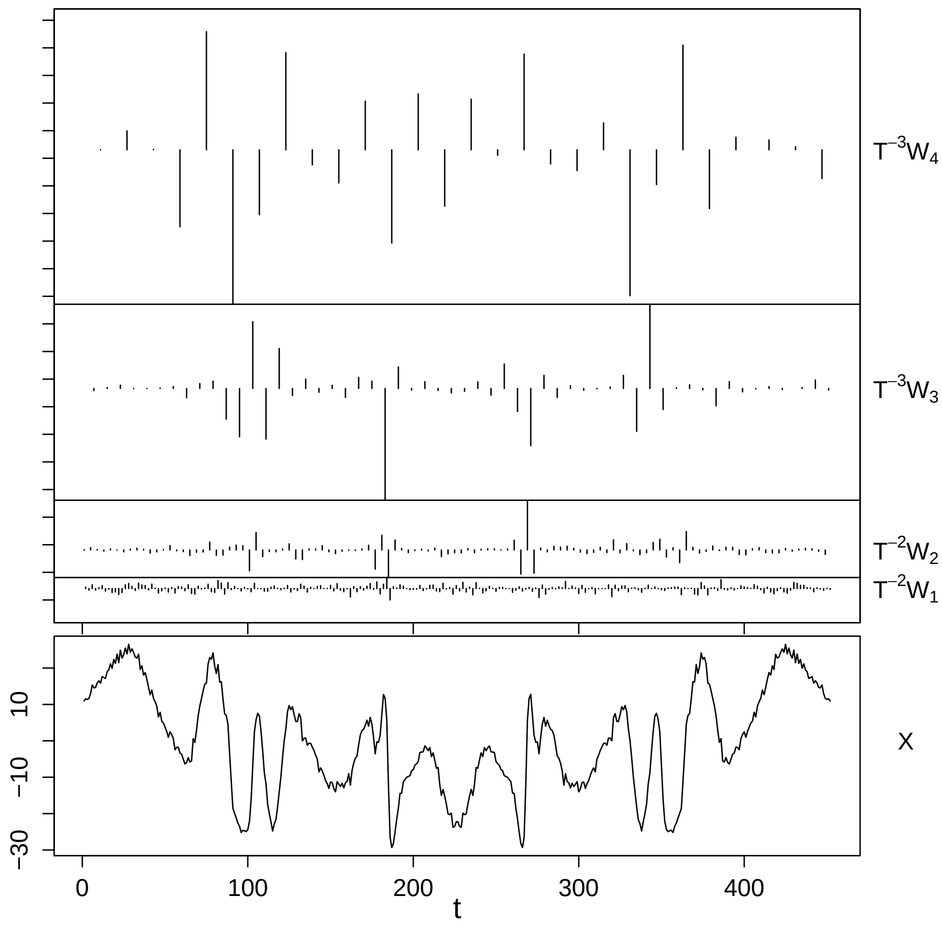


Figure 12. Example Vehicle Decomposition

Experiment 2 and the analyzed results present the possibility of using a smart phone as a type of scanning device. As scanning for activity has already been researched and proven highly accurate in numerous studies, this ability comes as no surprise. The ability to classify additional entities via such techniques presents numerous opportunities for future research.

VI. Conclusions

6.1 Entity Recognition

Analysis of the experiment output revealed the ability to accurately classify entities via sensor data gathered by smart phones. The ability to accurately classify entities has implications across a number of disciplines. The fidelity offered by the diverse set of sensors included with smart phones, as well as the current trend of adding additional sensors to smart phones, opens up an exciting world of classification via smart phone.

6.2 Implications

The ability to recognize entities with sensors that are readily available in smart phones opens up a number of possibilities, far too many to list exhaustively. Possible avenues for entity classification cover the gamut from a simple logging mechanism to detailed forensic analysis of a smart phone.

Smart phone users have access to apps that allow for recognition of a users activity. As noted, these apps allow a user to identify not just the presence of activity, but the form of activity, type of transportation, and with a large enough feature set, the location of the smart phone during the activity. Entity recognition could allow a user to identify microwave oven usage, time spent at a computer as compared to watching television, identification of a particular vehicle being driven, exposure to overly loud music, and many other scenarios.

Taking things a step further. It may be possible to identify that a smart phone user has entered a vehicle, and depending on smart phone carry location and magnetic signature, whether they are a driver or passenger. Using accelerometer data the motion of the vehicle could be analyzed and assessed. If for instance the algorithm determined a vehicle has been involved in an accident, it may be possible to alert first responders to the potential of a vehicle in distress.

Analysis of smart phones involved in house fires may reveal that there are detectable signals. With the inclusion of barometers and thermometers in smart phones, there is the possibility for data streams that could help alert first responders to the presence of an entity requiring attention.

From a different perspective, the ability to analyze entities from the point of view of the first responders may allow for the near instantaneous dispatch of additional assets. With the ability to analyze accelerative and decelerative patterns from a point of transportation, it should be possible to identify when a traffic officer gives chase. The same activity recognition algorithms could identify when an officer has to leave their vehicle, either to enter a new chase phase or issue a ticket. If an officer is put into a situation where they have to fire their sidearm, it may be possible due to a potential compression in the air surrounding the sidearm or via the microphone to determine the sidearm was discharged. This determination could be pushed immediately to dispatch, allowing for more rapid response to situations.

The ability to analyze a threat environment and feed information to dispatchers is not limited to police officers. In a combat environment, similar occurrences may also be detectable and thus able to be fed back to a command center for additional processing and/or action. The potential to help first responders, crisis responders, and combat personnel could prove helpful to commanders of all types.

The ability to scan an entity and determine which classification it fits into can also aid in threat detection. If a vehicle is known to produce certain effects on a sensor and it is producing effects that don't corroborate with expectations, there may be a need for further investigation. Sometimes the effects detectable by the eyes or a camera don't tell the whole story, a magnetic analysis may reveal the presence of anomalies.

Flocking observed from the behavior of large groups of people, combined with entity detection could be used to determine whether an active-shooter situation is taking place or not. Scattering and/or hunkering down could be used to determine an anomaly is present in the function of how people behave. Additional input from microphones and other sensors may aid in locating a perpetrator.

The ability to collect data from sensors and save the entities interacted with has the potential to analyze smart phones in a forensic manner. This could be done to prove timelines and whereabouts of a smart phone, and the associated user presumably, providing a signature of sorts for determining where and what a user was doing.

6.3 Further Research

The work performed in this thesis helped determine that data gathered from smart phone sensors was capable of being analyzed to accurately recognize the entities used as control variables. Some of the control variables were in categories disparate enough from one another that inter-category misclassification proved unlikely. However, at the subcategory level, between subwoofer frequencies and/or dB levels, the smart phone sensors were able to read environmental variables with enough fidelity to subcategorize the control variables highly accurately. In order to progress this research further, a number of issues must be studied further.

The first issue comes from the limited pool of entities studied in this thesis. The thesis proved that different entities of similar nature can be accurately identified with the correct set of attributes. Enlarging the pool of entities with a standard testing platform would allow for the expansion of entities recognizable by a smart phone. With enough entities, it may be possible to track someone's day in not just terms of activity, but also terms of interaction.

The second issue is the testing parameters. There are any number of experimental designs that are possible to implement when acquiring entity readings from smart phone sensors. A few of the more readily apparent are surface placement, smart phone mobility, and distances. The smart phone can be placed on any number of different surfaces, each having a different ability to vibrate and thus readily effecting accelerometer and gyroscope measurements. The smart phone can be placed in a manner that restricts movement through some hard attachment process or it can be laid flat on a surface that vibrates freely and thus may move the phone. Distances matter greatly when detecting the magnetic field generated by control variables. These are just three of the considerations that need to be addressed when designing an experiment.

The third issue has to do with the decision trees. The attributes were those identified in the literature review as working well for activity recognition. The attributes utilized worked for the entities chosen as control variables in the experiments discussed in this thesis. Other attributes not included in the decision trees discussed may be required to identify other entities. Additionally, it may be that wavelet decomposition when applied to the control variables in experiment 1 would work just as well as wavelet decomposition did for experiment 2.

The fourth issue is related to signatures. Each entity was captured and analyzed as a full signature after being trimmed by the windowing algorithm. This requires an identifiable sensing of start and stop points for the windowing algorithm in order to produce a signature for classification purposes. A process that captures a sampling for some time interval during an entities active phase would prove more useful than the requirement of a complete entity signature.

The fifth issue has to do with the windowing algorithm. The algorithm senses a start and stop point based off sensor data. This works for some entities, but not all. There are entities where a magnetic field may be experienced long before the accelerometer detects changes in gravity or the gyroscope detects torque on the cell phone. In the experiments discussed herein, the magnetometer was the source of trim points for all entities sans the two lowest dB level subwoofer. Figuring out how to tie the sensors together into a coherent windowing algorithm may be necessary if issue four above cannot be resolved.

The research accomplished in this thesis proved the ability to utilize the sensors embedded in smart phones in order to sense and classify entities external to the phone. The magnetometer, accelerometer, and gyroscope proved able to sense their respective environmental attributes at a resolution adequate to accurately identify several entities. These entities produced effects that were both unique and similar to one another, requiring attributes from multiple sensors in order to obtain the most accurate results. As such, a multimodal approach to sensor fusion was tested and validated, paving the way for further research.

Bibliography

1. Alvina Anjum and Muhammad U. Ilyas. Activity recognition using smartphone sensors. *2013 IEEE 10th Consumer Communications and Networking Conference (CCNC)*, pages 914–919, January 2013.

2. Christopher Barthold, Kalyan Pathapati Subbu, and Ram Dantu. Evaluation of gyroscope-embedded mobile phones. *2011 IEEE International Conference on Systems, Man, and Cybernetics*, pages 1632–1638, October 2011.

3. Tomas Brezmes, Juan-luis Gorricho, and Josep Cotrina. Activity Recognition from Accelerometer Data on a. pages 796–799, 2009.

4. A.J. Bernheim Brush, Kerry Hammil, Steven Levi, Amy K. Karlson, James Scott, Raman Sarin, Andy Jacobs, Barry Bond, Oscar Murillo, Galen Hunt, and Mike Sinclair. User experiences with activity-based navigation on mobile devices. *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services - MobileHCI '10*, page 73, 2010.

5. Michael J Caruso, Tamara Bratland, Carl H Smith, and Robert Schneider. A new perspective on magnetic field sensing. *SENSORS-PETERBOROUGH-*, 15:34–47, 1998.

6. Joshua J. Cogliati, Kurt W. Derr, and Jayson Wharton. Using CMOS Sensors in a Cellphone for Gamma Detection and Classification. pages 1–26, January 2014.

7. Electronic Compass. AK8963 3-axis Electronic Compass. pages 1–38, 2013.

8. Ingrid Daubechies. *Ten lectures on wavelets.* Society for Industrial and Applied Mathematics, 1992.

9. S Hochbaum Dorit. The multi-sensor nuclear threat detection problem. In *Operations Research and Cyber-Infrastructure*, pages 389–399. Springer, 2009.

10. Richard Duda, Peter Hart, and David Stork. *Pattern classification.* John Wiley & Sons, 2nd edition, 2000.

11. Matthew Faulkner, Robert Clayton, Thomas Heaton, K Mani Chandy, Monica Kohler, Julian Bunn, Richard Guy, Annie Liu, Michael Olson, MingHei Cheng, and Others. Community Sense and Response Systems: Your Phone as Quake Detector, 2013.

12. Matthew Faulkner, Michael Olson, Rishi Chandy, Jonathan Krause, K Mani Chandy, and Andreas Krause. The next big one: Detecting earthquakes and other rare events from community-based sensors. In *Information Processing in Sensor Networks (IPSN), 2011 10th International Conference on*, pages 13–24. IEEE, 2011.

13. Davide Figo, Pedro C. Diniz, Diogo R. Ferreira, and João M. P. Cardoso. Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Computing*, 14(7):645–662, March 2010.

14. P J Flynn. Data Clustering : A Review. 2000.

15. Auroop Ganguly, Joao Gama, Olufemi Omitaomu, Mohamed Gaber, and RAnga Vatsavai. *Knowledge discovery from sensor data.* Taylor & Francis, 2009.

16. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

17. Manhyung Han, La The Vinh, Young-Koo Lee, and Sungyoung Lee. Comprehensive Context Recognizer Based on Multimodal Sensors in a Smartphone. *Sensors*, 12(12):12588–12605, September 2012.

18. Samuli Hemminki, Petteri Nurmi, and Sasu Tarkoma. Accelerometer-Based Transportation Mode Detection on Smartphones Categories and Subject Descriptors.

19. Apiwat Henpraserttae, Surapa Thiemjarus, and Sanparith Marukatat. Accurate Activity Recognition Using a Mobile Phone Regardless of Device Orientation and Location. *2011 International Conference on Body Sensor Networks*, pages 41–46, May 2011.

20. Jeffrey Hightower, Roy Want, and Gaetano Borriello. SpotON: An indoor 3D location sensing technology based on RF signal strength. *UW CSE 00-02-02, University of Washington, Department of Computer Science and Engineering, Seattle, WA*, 1, 2000.

21. Dorit S Hochbaum and Barak Fishbain. Nuclear threat detection with mobile distributed sensor networks. *Annals of Operations Research*, 187(1):45–63, 2011.

22. Baik Hoh, Marco Gruteser, Ryan Herring, Jeff Ban, Daniel Work, Juan-Carlos Herrera, Alexandre M Bayen, Murali Annavaram, and Quinn Jacobson. Virtual trip lines for distributed privacy-preserving traffic monitoring. In *Proceedings of the 6th international conference on Mobile systems, applications, and services*, pages 15–28. ACM, 2008.

23. Barbara Hubbard. *The world according to wavelets: the story of a mathematical technique in the making.* A K Peters, Ltd., 2nd edition, 1998.

24. Mikkel Baun Kjæ rgaard, Martin Wirz, Daniel Roggen, and Gerhard Tr. Detecting Pedestrian Flocks by Fusion of Multi-Modal Sensors in Mobile Phones. 2012.

25. Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. Activity Recognition using Cell Phone Accelerometers. 12(2):74–82, 2010.

26. Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, 12(2):74–82, 2011.

27. Daniel Larose. *Data mining methods and models*. John Wiley & Sons, 2006.

28. Hong Lu, Jun Yang, Zhigang Liu, Nicholas D. Lane, Tanzeem Choudhury, and Andrew T. Campbell. The Jigsaw continuous sensing engine for mobile phone applications. *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems - SenSys '10*, page 71, 2010.

29. Steven A Macintyre. Magnetic field measurement. *ENG Net Base*, 2000.

30. Houghton Mifflin. *The American Heritage College Dictionary*. Houghton Mifflin, 4th edition, 2002.

31. Tom M Mitchell. Mining our reality. *Science*, 326(5960):1644–1645, 2009.

32. D. Mizell. Using gravity to estimate accelerometer orientation. *Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings.*, pages 252–253, 2003.

33. Anurak Poolsawat, Wasan Pattara-atikom, and Boonchai Ngamwongwattana. Acquiring road traffic information through mobile phones. *2008 8th International Conference on ITS Telecommunications*, (3):170–174, October 2008.

34. Nishkam Ravi, Nikhil Dandekar, Preetham Mysore, and Michael L Littman. Activity Recognition from Accelerometer Data.

35. Sasank Reddy, Min Mun, Jeff Burke, Deborah Estrin, Mark Hansen, and Mani Srivastava. Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks*, 6(2):1–27, February 2010.

36. S Sanfilippo. Hall probes: physics and application to magnetometry. *arXiv preprint arXiv:1103.1271*, 2011.

37. Corina Kim Schindhelm. Activity recognition and step detection with smartphones: Towards terminal based indoor positioning system. *2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC)*, pages 2454–2459, September 2012.

38. Arie Sheinker, Nizan Salomonski, Boris Ginzburg, A Shkalim, Lev Frumkis, and B Z Kaplan. Network of remote sensors for magnetic detection. In *Information Technology: Research and Education, 2006. ITRE'06. International Conference on*, pages 56–60. IEEE, 2006.

39. STMicroelectronics. LIS331DLH MEMS Accelerometer Data Sheet, 2009.

40. STMicroelectronics. L3G4200DH MEMS Gyroscope Data Sheet, 2010.

41. Lin Sun, Daqing Zhang, Bin Li, Bin Guo, and Shijian Li. Activity recognition on an accelerometer embedded mobile phone with varying positions and orientations. In *Ubiquitous intelligence and computing*, pages 548–562. Springer, 2010.

42. Luis Torgo. *Data mining with R: learning with case studies.* Chapman & Hall, 2010.

43. La The Vinh, Sungyoung Lee, Young-Tack Park, and Brian J. DAuriol. A novel feature selection method based on normalized mutual information. *Applied Intelligence*, 37(1):100–120, August 2011.

44. Shuangquan Wang, Canfeng Chen, and Jian Ma. Accelerometer Based Transportation Mode Recognition on Mobile Phones. *2010 Asia-Pacific Conference on Wearable Computing Systems*, pages 44–46, 2010.

45. Ian Witten, Eibe Frank, and Mark Hall. *Data mining: practical machine learning tools and techniques*. Elsevier, 3rd edition, 2011.

46. Changqing Zhou, Dan Frankowski, Pamela Ludford, Shashi Shekhar, and Loren Terveen. Discovering Personal Gazetteers : An Interactive Clustering Approach. 2004.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE (DD–MM–YYYY)	2. REPORT TYPE	3. DATES COVERED (From — To)
26–03–2015	Master's Thesis	Sept 2013 — Mar 2015

4. TITLE AND SUBTITLE

Entity Recognition Via Multimodal Sensor Fusion With Smart Phones

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Nagy, John E. Capt, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENG-MS-15-M-023

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Intentionally Left Blank

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This thesis serves as an exploration that takes the sensors within a cell phone beyond the current state of recognition activities. Current state of the art sensor recognition processes tend to focus on recognizing user activity. Utilizing the same sensors available for user activity classification, this thesis validates the ability to gather data about entities separate from the user carrying the smart phone. With the ability to sense entities, the ability to recognize and classify a multitude of items, situations, and phenomena opens a new realm of possibilities for how devices perceive and react to their environment.

15. SUBJECT TERMS

Smart Phone, Sensors, Magnetometer, Accelerometer, Gyroscope, Entity Recognition

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Dr. K. M. Hopkinson, AFIT/ENG
					19b. TELEPHONE NUMBER (include area code)
U	U	U	U	122	(937) 255-3636, x4579; kenneth.hopkinson@afit.edu

Standard Form 298 (Rev. 8–98)
Prescribed by ANSI Std. Z39.18